Article

Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications

Yufeng Qian[1] 🆔

Abstract

The effectiveness of generative AI tools in education depends largely on prompt engineering—the practice of designing inputs and interactions that guide AI systems to produce relevant, high-quality outputs. This systematic literature review examines empirical studies published since the release of ChatGPT in late 2022, identifying two broad approaches of prompting strategies: technique-based, which targets specific learning goals, and process-based, which supports cognitive engagement and collaborative thinking with AI. The review identifies key educational applications of prompt engineering, notably in two overarching areas: critical skills development and the automation of educational functions. It also highlights emerging trends, such as the integration of multimodal AI and the growing influence of advanced AI reasoning capabilities. By mapping this evolving landscape, the findings provide a foundational understanding of prompt engineering as both a technical skill and a pedagogical strategy in AI-supported learning environments.

Keywords

artificial intelligence, prompt engineering, educational applications, systematic review

[1]College of Human Sciences and Education, Louisiana State University, Baton Rouge, LA, USA

Corresponding Author:
Yufeng Qian, Lutrill & Pearl Payne School of Education, College of Human Sciences & Education, Louisiana State University, 221 Peabody Hall, Baton Rouge, LA 70803, USA.
Email: jqian@lsu.edu

Introduction

Generative AI programs, such as OpenAI's ChatGPT and DALL-E, Google's Gemini, Anthropic's Claude, and the most recent competitors - DeepSeek and Grok - have rapidly emerged as transformative educational tools. However, their effectiveness depends significantly on prompt engineering—designing effective inputs and interactions that guide AI systems to produce relevant, high-quality responses (Cain, 2024; Park & Choo, 2024).

Educators and researchers are actively exploring a range of prompt engineering techniques to improve the quality and relevance of AI-generated outputs. One widely used approach is few-shot prompting, which involves supplying the AI with a few examples to guide its responses, helping it better understand the desired output format or reasoning style (Schick & Schütze, 2022). Another effective method is chain-of-thought prompting, which encourages the AI to articulate its reasoning step by step. This approach is especially valuable for tasks that demand logic, problem-solving, or critical thinking (Wei et al., 2022). In addition to these techniques, the CLEAR framework developed by Lo (2023) offers a set of criteria for crafting and evaluating effective prompts. CLEAR stands for Concise, Logical, Explicit, Adaptive, and Reflective—qualities that promote clarity, structure, transparency, learner-centeredness, and critical engagement.

Despite this progress, the field of prompt engineering is still in its infancy. Like many emerging technologies in the history of educational technology, early adoption tends to rely on trial-and-error methods, with educators testing what works in real-world contexts (Selwyn, 2022). This exploratory phase is not only expected but necessary. It lays the groundwork for future research, theory-building, and professional practice. This same process of exploration and refinement is now unfolding in prompt engineering, which is poised to become a core competency for teaching and learning in AI-enhanced environments.

Background

Prompt engineering refers to the practice of carefully designing and refining prompts—specific instructions or queries—to guide AI systems, particularly large language models, toward generating desired responses (Brown et al., 2020; Liu et al., 2021). While the concept has gained prominence with the advent of advanced LLMs, its roots trace back to earlier natural language processing (NLP) tasks aimed at enhancing the accuracy and relevance of automated responses (Radford et al., 2019). The release of OpenAI's ChatGPT in late 2022 significantly elevated the importance of prompt engineering, underscoring its role in maximizing the performance and utility of generative AI tools (OpenAI, 2022).

Existing systematic reviews on prompt engineering in education have primarily focused on specific domains, offering valuable but fragmented insights rather than a comprehensive understanding across educational contexts. For example, Chen, Liu et al. (2024) focused on K–12 STEM education, exploring how tailored prompts

influence student engagement and learning outcomes in science and mathematics. Sahoo et al. (2024) provided a broader technical review of prompt engineering strategies, though with an emphasis on computational mechanisms rather than pedagogical use, limiting its relevance to classroom practice. More recently, Lee and Palmer (2025) categorized prompt engineering studies in higher education into themes such as skills, administration, creativity, frameworks, and shots. Their review, which included both empirical and conceptual work, illustrated the diversity of prompt engineering use cases.

These reviews highlight the growing interest in prompt engineering while revealing a gap in synthesizing findings across educational levels and disciplines. This gap calls for a systematic investigation into the evolving landscape of prompt engineering: How is prompt engineering being implemented in educational settings? What emerging patterns and approaches are observable in the early stages of its use? Furthermore, how do prompt engineering strategies facilitate specific educational uses within these contexts? This study addresses these questions through a systematic review of empirical studies on the educational use of prompt engineering in the first two years following the release of generative AI tools like ChatGPT. Given that the field of prompt engineering is still in its infancy, the study aims to establish an initial framework to guide educators and researchers in the effective use of prompt engineering techniques.

The purpose of this review is to examine the early applications of prompt engineering in education and to begin constructing a foundational framework for its use. To guide this investigation, the study is structured around the following research questions:

(1) What prompt engineering approaches have been implemented in educational settings during the initial two years of adoption?
(2) In what ways do prompt engineering approaches support different educational applications?

Method

This systematic review used three widely recognized databases—ERIC, Web of Science, and ScienceDirect —that are commonly used for systematic reviews in education and technology. ERIC is particularly known for its extensive coverage of education-related research, while Web of Science and ScienceDirect provide a broad range of interdisciplinary studies, including technology and education. In addition, to ensure the review captured emerging trends and publications not yet indexed by traditional databases, a recent AI-powered scholar database, SciSpace, was included. SciSpace was specifically chosen to broaden the search scope and identify publications from newer or interdisciplinary journals that may not be fully covered by the other three databases. However, to maintain consistency with established academic standards and focus on peer-reviewed research, preprints from SciSpace were excluded from the review.

The search strings used for this review were intentionally broad to ensure a wider range of relevant studies. The keywords "prompt engineering" and "education" were chosen for their ability to capture a broad spectrum of literature. The search was conducted across the four databases mentioned, and the review focused on publications from January 2023 to February 2025, including early online releases. Studies were excluded if they were not in English or if full-text access was unavailable, ensuring that only accessible and complete studies were included in the analysis.

The systematic review followed a rigorous screening process following the PRISMA framework (Page et al., 2021). Initially, 243 records were identified from four databases: ERIC (18), ScienceDirect (33), Web of Science (92), and SciSpace (100). Notably, the paid version of SciSpace can generate up to 100 sources, contributing significantly to the initial pool. After removing 28 duplicate records, 215 unique records were screened. All were sought for retrieval, but five studies could not be accessed in full text. The remaining 210 reports were assessed for eligibility, leading to the exclusion of 160 studies due to reasons such as not focusing on educational applications of prompt engineering ($n = 91$), being technical or programming-related ($n = 22$), editorial articles ($n = 5$), book chapters ($n = 3$), conference synopsis ($n = 12$), SciSpace preprints ($n = 8$), conceptual frameworks ($n = 16$), and review articles ($n = 3$). Ultimately, 50 studies were included in the final review (see Figure 1).

Descriptive metadata, including year of publication, the country of the first author, educational level, and academic discipline, was initially collected for each article. This metadata facilitated the capture of both geographical and academic discipline distributions within the field, helping to situate the study's broader context. To address the study's central inquiry—what prompt engineering methods are used, for which educational applications, and how they support those applications—the three levels of grounded theory coding (Strauss & Corbin, 1998) were employed. Initially, open coding was used, involving a line-by-line analysis of all articles to identify initial concepts without imposing any prior frameworks. This phase allowed for an inductive approach where raw data from the articles were broken down into discrete units, capturing key ideas as they emerged from the content. The purpose was to ensure that no aspect of the data was overlooked, enabling the generation of preliminary codes that reflected the diverse nature of the prompts being analyzed. Following this, focused coding was employed to refine and consolidate the initial codes. The analysis identified patterns in the data, grouping the open codes into broader categories that reflected more meaningful distinctions. For example, prompts that were initially categorized as "zero-shot prompt," "direct prompt," "open prompt," and "creative prompt" were consolidated under the broader category of unstructured prompting, characterized by minimal guidance or constraints. On the other hand, prompts involving predefined steps or explicit structures were grouped under structured prompting, reflecting a key distinction in the design of the prompts and the guidance provided to learners. This phase also helped simplify the data, making it more manageable for subsequent analysis while still maintaining the richness of the original insights. Finally, in the theoretical
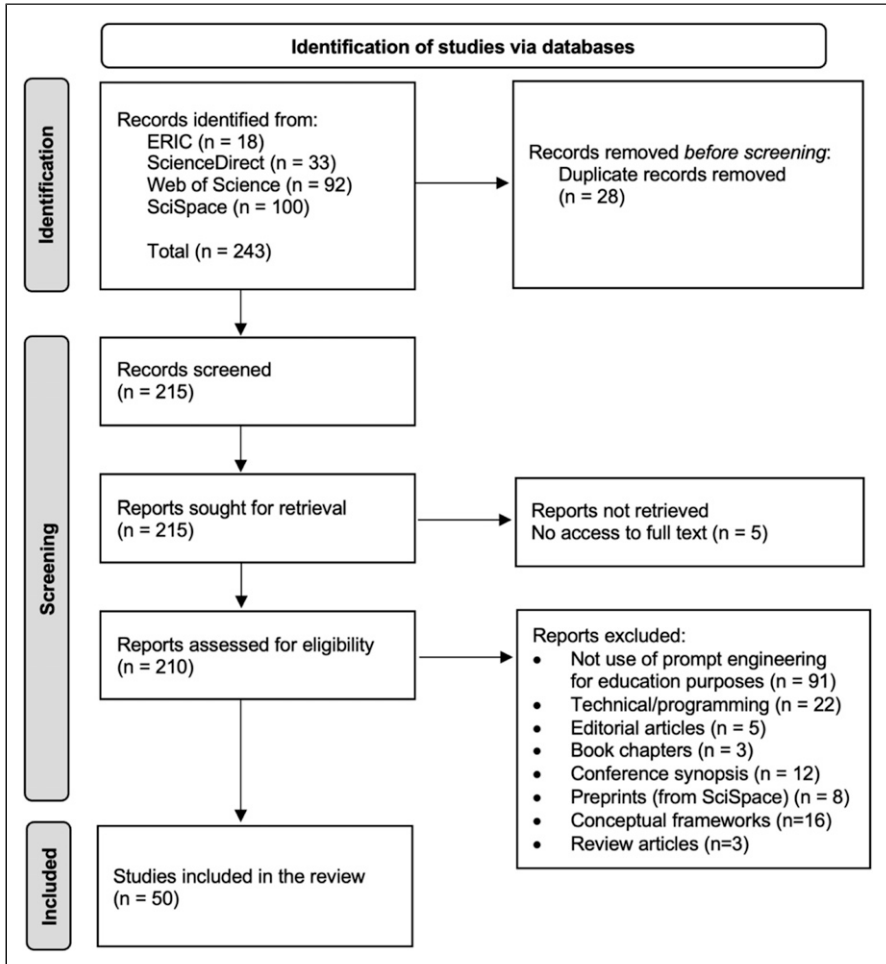
Figure 1. PRISMA Flow Chart of Article Identification and Screening (Page et al., 2021)

coding phase, these synthesized categories were linked to the study's overarching research questions, helping to build an initial conceptual framework. For instance, the theme of "critical skills development" emerged from the grouping of codes like "prompt literacy," "scientific reasoning," and "creativity." This theme was particularly relevant to Research Question 2, which examined the educational applications of the prompting approaches.

NVivo 14 was employed as an additional researcher in the traditional coding and thematic analysis process, providing an AI-powered approach to generate preliminary coding suggestions. These initial codes were reviewed, refined, and expanded upon by the author, ensuring that the AI's input aligned with the research

objectives. After the author manually coded the full dataset, a 10% subset was cross-verified using NVivo's Coding Comparison Query tool to assess inter-rater reliability. This step was crucial for ensuring the consistency of the coding process and the alignment of the AI's suggestions with the author's interpretations. Any discrepancies between the author's and AI-generated codes were addressed through iterative calibration, during which the codebook definitions were refined, and the AI was retrained based on the updated criteria. Following three rounds of revision, full consensus was achieved between the human coder and the AI, ensuring a high level of reliability and consistency in both the code definitions and thematic groupings. This rigorous process of validation and refinement supported the robustness of the final analysis and the alignment of the AI-generated codes with the research questions.

Results and Discussion

The results and discussion of the systematic review are structured into three main sections: Geographical and Disciplinary Distribution, Research Question 1, and Research Question 2. The first section provides an overview of the distribution of studies based on geography and academic discipline during the first two years following the release of ChatGPT. The Research Question 1 section delves into the analysis and categorization of specific prompting approaches used across the studies. Lastly, the Research Question 2 section examines the educational applications of prompting, focusing on how these methods are employed to support learning. Together, these sections offer valuable insights into the evolving role of prompt engineering and its impact in education.

### Geographical and Disciplinary Distribution

The distribution of first authors by country reveals a strong representation from the United States, which contributes 50% of the articles, with 25 studies, followed by South Korea with 6. China (including Hong Kong) produced 5 studies, while Singapore contributed 3. Australia and Germany each accounted for 2 studies. The remaining countries, including Austria, Italy, Kuwait, Portugal, Spain, Switzerland, and Vietnam, were represented by a single study each. This distribution highlights the global interest in prompt engineering within education, with a particularly strong concentration of research emerging from the United States and East Asia. See Figure 2.

Figure 3 illustrates the disciplinary distribution of studies included in the review. Medical education leads with 24 studies, underscoring its early and substantial adoption of prompt engineering for clinical training, assessment, and decision support. STEM disciplines collectively account for 11 studies across chemistry, mathematics, and physics, showcasing the diverse applications of prompt engineering in scientific and technical contexts. The remaining 15 studies are distributed across the social sciences and humanities, including language learning (such as
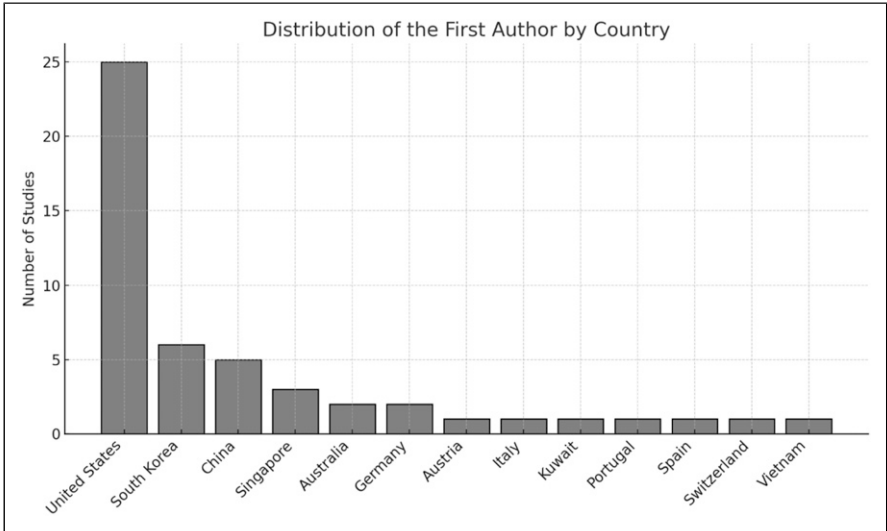
Figure 2. Distribution of Country (by First Authors)

ESL/EFL and academic writing), business, teacher education, general education, arts and design, and journalism. This distribution indicates that, within the first two years since the release of ChatGPT, prompt engineering has gained the most traction in medical and STEM-related fields.
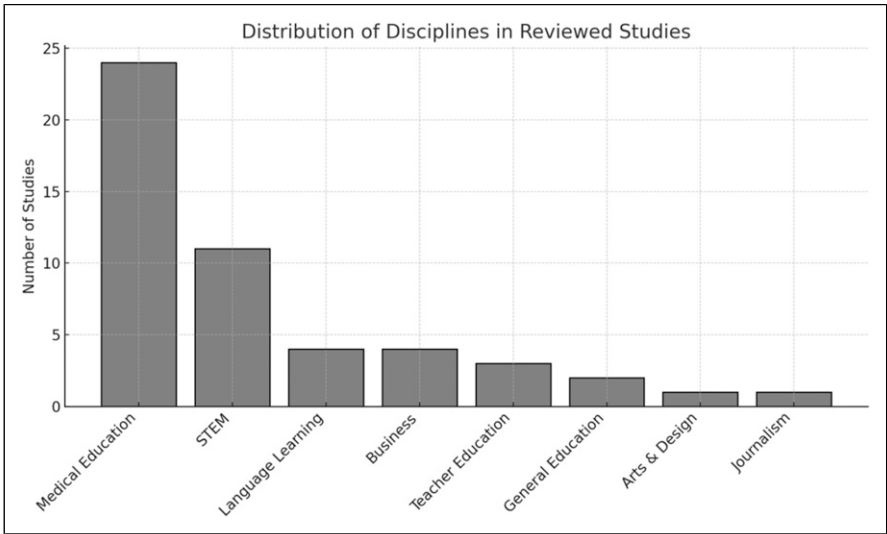
Figure 3. Distribution of Academic Disciplines

Research Question 1: What Prompt Engineering Approaches Have Been Implemented in Educational Settings During the Initial Years of Adoption?

The coding of prompting approaches yielded 57 distinct codes, including techniques such as zero-shot, one-shot, few-shot, chain-of-thought, retrieval-augmented generation (RAG), iterative refinement, human-in-the-loop, and goal setting–refining–revising. Through focused coding, these were grouped into two overarching categories: (1) techniques and (2) processes. Techniques refer to specific prompting methods like few-shot prompting or chain-of-thought reasoning, while processes encompass broader, step-by-step workflows that often integrate multiple techniques.

Of the 50 studies reviewed, 23 studies focused solely on techniques, reflecting a focused and valuable exploration of prompt design. 27 addressed both techniques and processes, emphasizing how strategic workflows and tool iterations shape prompt effectiveness. Unlike earlier reviews that primarily centered on prompting techniques, this study uniquely identifies and distinguishes between technique-focused and process-integrated approaches, offering a more comprehensive understanding of how prompt engineering is being applied in educational contexts. Further details on these two categories are discussed in the sections below.

Technique-Based Prompting. This section highlights five key categories of prompting techniques identified in the review, each offering distinct approaches for guiding AI models in educational settings, from structured prompting with predefined constraints to unstructured prompting that allows for greater flexibility. Figure 4 lists each category within specific strategies used in the reviewed studies.

1. Structured Prompting. Structured prompting emerged as the most frequently utilized strategy, appearing in 22 studies (e.g., Abdullahi et al., 2024; Lee, Jung, et al., 2025; Wan & Chen, 2024). They guide AI model behavior using predefined structures, implemented through three distinct strategies: examples, outcomes, and assessment.

Contextual example-based prompts provide example-based prompts guide the model by providing explicit examples, helping it understand the expected behavior. In the *one-shot* approach, a single example is used to demonstrate the desired response. For example, in Wan and Chen's (2024) study, GPT-3.5 was given an example of feedback on a physics conceptual question, which enabled the model to replicate a similar style and content of effective feedback in subsequent responses. On the other hand, the *few-shot* method, which uses 3–5 examples, further fine-tunes the model's understanding. Abdullahi et al. (2024) applied this approach with GPT-4 to diagnose rare diseases, presenting a small set of medical cases to guide the model's diagnostic conclusions. One-shot learning helps the model refine its outputs by providing just enough context to generate relevant responses, while few-shot learning is particularly effective for tasks that require more context than a single example but still benefit from a limited number of examples.
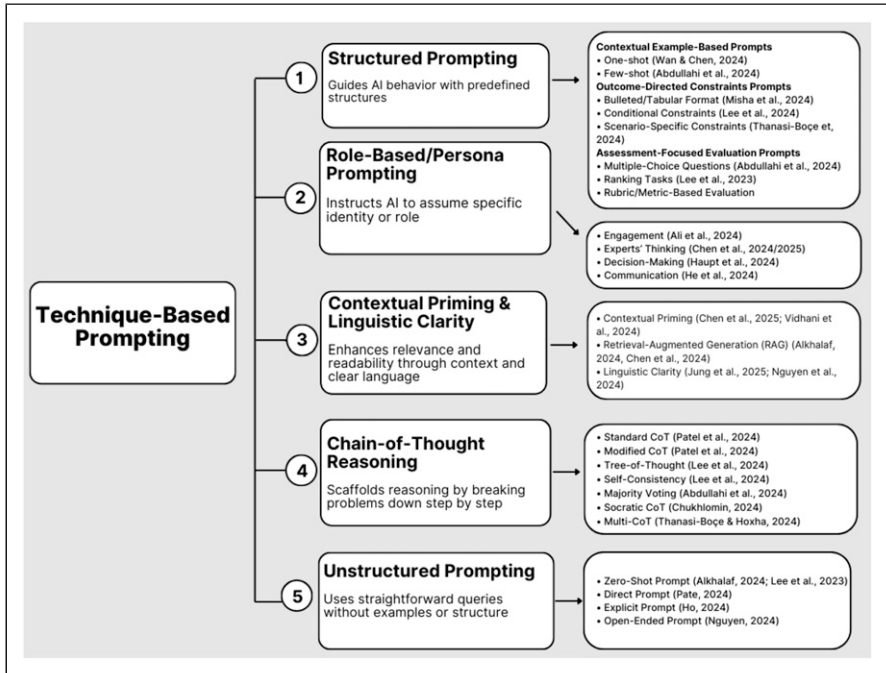
Figure 4. Technique-Based Prompting

Outcome-directed constraints prompts steer the model's responses by setting predefined structures or constraints that narrow the range of possible outcomes. For example, Mishra et al. (2024) used a *bulleted format* in GPT-4 to simplify complex cardiovascular disease information, improving readability and comprehension, especially for technical content. *Conditional constraints* limit the model's responses to specific conditions or steps. In Lee, Kim et al. (2024), GPT-4 was guided through reasoning steps in educational contexts, particularly in mathematical problem-solving, ensuring logical progression. *Scenario-specific constraints* focus the model's output on specific contexts. Thanasi-Boçe and Hoxha (2024) applied structured prompts in entrepreneurial education, directing GPT-4 to generate business plans under constraints like limited budgets or timeframes, ensuring practical and relevant solutions. By defining parameters such as market size or target demographic, these prompts grounded the AI's output in realistic scenarios. The outcome-directed prompting ensures that AI produces results in the expected formats, steps, or scenarios.

In addition to providing examples and setting expected outcomes, the third strategy in structured prompting focuses on assessment. *Assessment-oriented prompts* are designed to evaluate the accuracy, completeness, and relevance of the model's responses using predefined criteria (Fernández et al., 2024). Abdullahi et al. (2024) assessed GPT-4's diagnostic accuracy using *multiple-choice questions*,

asking the model to select the most likely diagnosis from options based on a medical case. This method is efficient for tasks with binary responses. Similarly, Lee et al. (2023) used *ranking tasks* to assess the model's ability to prioritize answers. *Rubric- or metric-based evaluations* further refine assessments using predefined criteria. For instance, Wan and Chen (2024) evaluated GPT-3.5's feedback on student responses using a detailed rubric, ensuring consistency. Thanasi-Boçe and Hoxha (2024) applied a similar approach to evaluate GPT-4's business plans, while Lee, Latif et al. (2024) used rubrics for student science assessments, standardizing evaluations for reliable and relevant results. Together, these strategies automate assessments, ensuring more consistent, reliable, and aligned outputs.

2. Role-Based/Persona Prompting. Used in 12 studies, role-based prompting is a widely used technique in prompt engineering, where the AI model is instructed to assume a specific identity or professional persona, such as a teacher, physician, or policy analyst (Chen & Makmur, 2024; Haupt et al., 2024; He et al., 2024). This strategy is particularly effective in fields that require discipline-specific language, tone, and decision-making processes. By clearly framing the model's role, the technique helps elicit responses that align with domain expectations, making the model's outputs more relevant and trustworthy.

One prominent application is in supporting learner engagement through pedagogical identities. In K–12 education, role-based prompting transforms AI into a peer tutor, inquiry guide, or mentor. For example, when acting as a mentor, the AI encourages metacognitive reflection through reflective questions and feedback, mimicking effective teaching strategies. This approach enables AI to offer personalized, dialogic learning experiences that are aligned with instructional goals and students' needs (Ali et al., 2024).

Another key use of role-based prompting is in professional settings, particularly in clinical contexts. Here, AI is tasked with acting as a healthcare professional—such as an ophthalmologist—able to explain complex conditions and procedures with empathy and precision. Role-based prompts ensure that the AI adopts an appropriate tone, uses discipline-specific language, and presents information in ways that meet professional standards (Chen & Makmur, 2024; Chen, Reddy et al., 2024). This persona-based approach is also crucial for improving accessibility and understanding, especially when specialized knowledge must be communicated to a lay audience. For example, when instructed to act as a health educator, AI can simplify medical jargon, providing clearer, context-sensitive explanations and enhancing safety in patient communications (He et al., 2024).

However, role-based prompting also presents ethical risks. Studies examining misinformation detection found that assigning models social identities—such as political or religious affiliations—can influence the AI's judgment, reducing classification accuracy (Haupt et al., 2024). This highlights the cognitive bias that can arise when AI is prompted to simulate subjective viewpoints, underscoring the need for careful design and ethical oversight.

In sum, role-based/persona prompting allows AI to mirror human communication styles, goals, and responsibilities across diverse contexts. When applied

thoughtfully, it helps AI align with audience expectations and domain-specific standards. However, as demonstrated in misinformation detection studies, it is crucial to apply this technique with awareness of its potential to introduce bias, particularly in high-stakes or value-laden scenarios.

3. Contextual Priming and Linguistic Clarity. As the third major category in the technique-based approach, contextual priming and linguistic clarity appeared in 11 studies (e.g., Ellison et al., 2025; Jung et al., 2025). These studies underscore the importance of embedding contextual cues and linguistic scaffolds into prompt design to enhance the relevance, clarity, and appropriateness of AI-generated responses. By specifying tone, simplifying language, and incorporating metadata, such as target grade level, domain knowledge, or user expertise, researchers were able to guide language models to produce more precise and audience-aligned outputs.

Contextual priming involves enriching prompts with background information, situational details, or domain-specific language to help the model simulate situational awareness. This can improve the accuracy, coherence, and practical relevance of outputs. For example, in ophthalmology education, prompts that included patient symptoms and references to specific retinal conditions enabled ChatGPT to deliver more precise and tailored medical explanations (Jung et al., 2025). Similarly, in chemistry education, explicitly stating technical terms or clarifying symbols helped guide the model toward scientifically accurate responses (Vidhani & Mariappan, 2024), avoiding ambiguity that often arises from general prompts.

One particularly impactful strategy within this theme is *Retrieval-Augmented Generation (RAG)*—a method that supplements prompts with external information retrieved from curated sources such as textbooks, databases, or health records. While RAG is typically associated with architectural enhancements, it closely intersects with prompt engineering by embedding relevant, task-specific content into the model's context window before generation. For instance, in medical settings, combining contextual priming with retrieval methods improved the accuracy and depth of responses to patient questions (Alkhalaf et al., 2024; Chen, Reddy et al., 2024). These techniques demonstrate how prompts, when designed with dynamically retrieved knowledge, can elevate general-purpose models into specialized assistants without retraining.

Complementing contextual cues is the technique of *linguistic clarity*, which focuses on minimizing ambiguity and optimizing readability. Studies show that clear, direct language in prompts enhances the interpretability and accuracy of AI responses, especially when communicating with non-expert audiences. According to Nguyen et al. (2024), simplifying the language of medical prompts—such as specifying a sixth-grade reading level—led to responses that were significantly more accessible to patients with limited health literacy. Similarly, Jung et al. (2025) found that tailoring prompts to emphasize readable structure and vocabulary improved both comprehension and user engagement in patient education materials.

As the third major category of technique-based prompting approach, contextual priming and linguistic clarity involve integrating background knowledge, using retrieval-enhanced inputs, and ensuring precise language. These strategies help AI

models generate responses that are not only accurate and well-informed but also audience-aware and accessible. It is recommended to combine contextual and linguistic techniques with structured prompts and role/persona-based prompting to enhance the relevance and impact of AI-generated content.

4. Chain-of-Thought Reasoning. Chain-of-Thought reasoning (CoT) prompting, featured in 10 studies (e.g., Chan et al., 2025; He et al., 2024; Lee, Teo et al., 2024), guides the model through a step-by-step reasoning process, making it particularly useful in fields like mathematics, clinical decision-making, and automated scoring. The most common approach is standard CoT, where the model is explicitly instructed to "think step by step." Patil et al. (2024) tested this in clinical education, prompting GPT-3.5 to solve USMLE-style clinical questions. While the improvement over direct prompting was modest (CoT scored 62.8% vs. 61.7%), it enhanced the transparency of the model's decision-making, a key benefit in medical education.

To address the limitations of basic CoT, researchers explored *modified Chain-of-Thought (mCoT)* strategies. For example, Patil et al. (2024) introduced scaffolding to guide the model through intermediate steps, but mCoT underperformed (scoring 57.4%) compared to standard CoT, suggesting that too much structure could restrict the model's reasoning flexibility. In mathematics education, more advanced variants such as *Tree-of-Thought (ToT)* and *Self-Consistency* were tested by Lee, Teo et al. (2024). ToT prompted the model to generate multiple solution paths, while self-consistency involved selecting the most consistent response from several CoT outputs. Although these methods improved interpretability and rubric-aligned scoring, they did not surpass human graders in complex reasoning tasks.

Further innovations in CoT include *Majority Voting*, as demonstrated by Abdullahi et al. (2024) in medical diagnostics. In this approach, multiple reasoning paths were generated, and the most frequent conclusion was selected as the final diagnosis, enhancing both accuracy and trust. *Socratic Prompting*, explored by Chukhlomin (2024), involved a conversational approach where the model iteratively questioned itself, fostering metacognitive reflection. This dialogic format proved valuable for both learners and instructional designers by encouraging reflective engagement. Lastly, *multi-CoT reasoning* was employed in entrepreneurship education by Thanasi-Boçe and Hoxha (2024), where the model was guided through decision points in business ideation, simulating real-world critical thinking. These CoT variants help improve the quality and reliability of AI-generated responses across education, healthcare, and entrepreneurship, demonstrating the importance of thoughtful prompt design.

Chain-of-Thought (CoT) prompting has emerged as a pivotal technique for enabling large language models to perform logical, multi-step reasoning. The above review highlights its key variants, such as modified CoT (m-CoT), Tree-of-Thought (ToT), Majority Voting, and Socratic dialogue. Collectively, these approaches underscore the role of strategic prompt design in improving the quality and reliability of AI-generated outputs across diverse domains, including education (Lee, Teo et al., 2024), healthcare (Patil et al., 2024), and entrepreneurship (Thanasi-Boçe & Hoxha, 2024), reflecting their wide applicability and impact.

5. *Unstructured Prompting.* Unstructured prompts are those that do not rely on formatted examples, predefined scaffolds, or step-by-step reasoning instructions. Instead, they allow the AI model to generate responses freely, drawing solely on its pretraining and the inherent openness or directness of the prompt. While less mentioned in the studies because the studies examined more complex prompt strategies, unstructured prompting is a standard method that involves straightforward queries to AI, allowing it to generate responses based on the input. It is favored for its simplicity and speed in obtaining answers (Abdullahi et al., 2024).

In some studies, what is referred to as *Direct Prompting* (Patil et al., 2024) is also labeled as *Zero-Shot Prompting* (e.g., Alkhalaf et al., 2024). In a zero-shot scenario, the model is expected to respond to a task without having seen any prior examples. This setup challenges the model to apply its learned knowledge independently (Sivarajkumar et al., 2024). For instance, in a study by Lee et al. (2023), GPT-4 was tasked with scoring open-ended mathematics responses in a zero-shot context. The absence of examples required the model to rely entirely on its language understanding to evaluate student answers without the aid of templates or rubrics. This approach is especially valuable for assessing a model's generalized reasoning ability across diverse topics.

A related strategy is the use of *Explicit Prompts* (Ho, 2024), which are characterized by their clarity and directness. These prompts clearly state the task without ambiguity—for example: *"What are the best practices for improving customer service?"* Their straightforward nature makes them particularly effective in early testing phases, where simplicity and interpretability are key.

In other cases, such as in Nguyen (2024), these prompts are described as *Open-Ended Prompts*. This variant emphasizes flexibility, inviting the model to generate expansive responses that may vary in content, style, and perspective—ideal for exploratory tasks or creative reasoning. Another type is *Creative Prompts* (Ho, 2024). This type of prompt offers the most flexibility, encouraging innovative and out-of-the-box thinking. Creative prompts might include phrases like "suggest innovative ways to engage customers," which allows the AI to generate unique and diverse ideas.

The five categories of technique-based prompting approaches presented above offer a comprehensive yet practical framework that educators and students can readily apply. For instance, role-based or persona prompting is particularly well-suited for scenario-based learning, allowing AI to simulate authentic roles like mentors, doctors, or policymakers. For tasks such as essay writing, educators may choose to upload a rubric or evaluation metric to guide students to work with AI in producing responses aligned with assessment criteria.

It is important to note that these categories often intersect in practice. Many studies combined multiple techniques to optimize performance. For example, in Chan et al.'s (2025) study, Chain-of-Thought prompting was integrated with few-shot learning, providing the model with a sample question-answer pair to scaffold step-by-step reasoning. This blending of methods illustrates the flexibility of prompt design and the potential for combining strategies to enhance reasoning, clarity, and alignment with user goals.
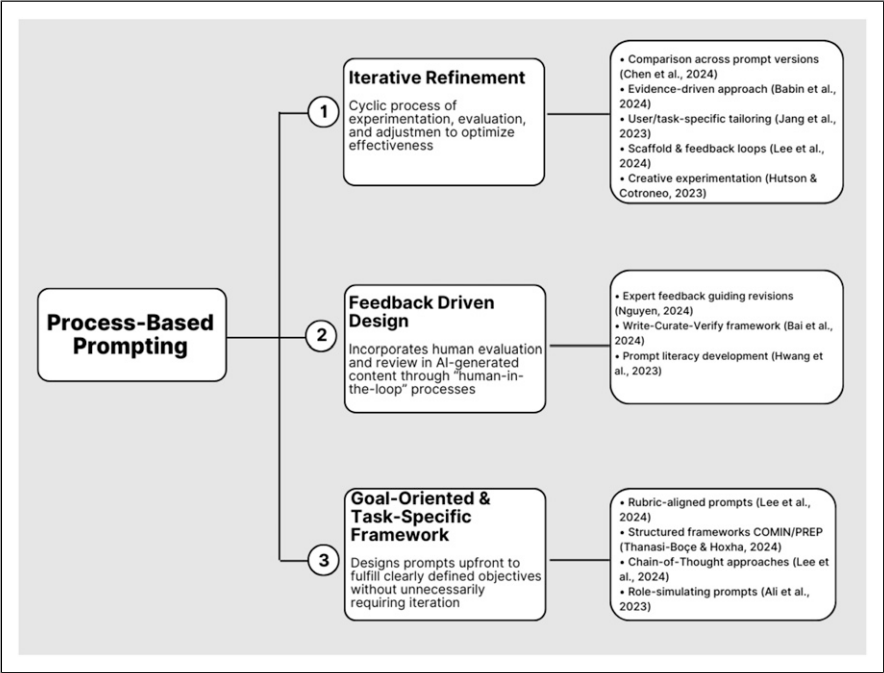
Figure 5. Process-Based Prompting

Process-Based Prompting. This section highlights key themes from studies adopting a "process-oriented" approach to prompt engineering, which focuses on structured workflows, iterative refinement, and feedback-driven methods to optimize generative AI tools. Three primary process-based strategies are explored: Iterative Refinement, Feedback-Driven Design, and Goal-Oriented & Task-Specific Framework, with an overview of these approaches provided in Figure 5.

1. Iterative Refinement. Iterative refinement involves a cyclic process of experimentation, evaluation, and adjustment aimed at optimizing prompts for specific contexts. A key feature is comparing different prompt versions. For example, Chen, Reddy et al. (2024) tested role-based framing and Retrieval-Augmented Generation (RAG) in clinical scenarios, refining prompts to improve accuracy and empathy. Similarly, Babin and Coberly (2024) adjusted the prompt structure based on expert consensus to better align with domain-specific needs. Iterative refinement also includes user-specific tailoring, achieved through scaffolding and feedback loops. Jang et al. (2024) demonstrated this with ChatGPT in K–12 education, where students iteratively refined prompts to improve their argumentative writing. In creative fields, Hutson and Cotroneo (2023) documented iterative experimentation by art and design students using DALL-E 2, refining prompts for more nuanced visual outcomes. This process of ongoing calibration continuously improves prompt

effectiveness, responding to user feedback, task demands, and performance outcomes (Venerito et al., 2024).

2. Feedback-Driven Design. Feedback-driven design emphasizes a "human-in-the-loop" approach, where human evaluators assess, score, or review AI-generated content to improve the model's outputs. Several studies incorporated user or expert feedback at various stages of prompt development. Nguyen et al. (2023) used persona-based evaluations to guide prompt revisions, while Bai et al. (2024) employed a Write–Curate–Verify framework in a computational thinking course, allowing teachers to refine scenarios using rubrics. In Hwang et al. (2023), feedback helped students improve their text-to-image prompts, fostering prompt literacy in language learners. This approach underscores that feedback is not just evaluative but generative, helping align AI outputs with learning goals and improving the quality and clarity of AI-generated content.

3. Goal-Oriented & Task-Specific Framework. Goal-oriented and task-specific frameworks focus on designing prompts to fulfill clearly defined instructional or functional objectives. These frameworks align prompt structures with desired outcomes, such as eliciting student reasoning or performing a diagnostic task. Lee, Kim et al. (2024) used goal-specific prompts to evaluate elementary students' responses in geometry, designing prompts that mirrored assessment structures and embedded rubric-aligned scoring. Thanasi-Boçe and Hoxha (2024) introduced structured frameworks for entrepreneurship education, incorporating stages like task definition and validation to ensure pedagogical alignment. Similarly, Ali et al. (2023) designed task-specific prompts for the TeacherGAIA chatbot, simulating teacher roles to foster metacognitive behaviors in K–12 students. These frameworks emphasize a proactive approach to prompt design, ensuring that AI behavior aligns precisely with educational goals from the outset.

The three process-based prompting approaches discussed above—iterative refinement, feedback-driven design, and goal-oriented and task-specific frameworks—are particularly well-suited for learning tasks that demand precision, adaptability, and alignment with clearly defined instructional objectives. Notably, these approaches are rarely implemented in isolation. Instead, they are most effective when integrated with technique-based prompting methods such as structured prompting, role assignment, contextual priming, and linguistic clarity, as consistently demonstrated across the studies reviewed. As extant systematic reviews have largely focused on techniques, the findings on process-based promoting highlight a promising direction for future research and development in prompt engineering.

Research Question 2: In What Ways Do Prompt Engineering Approaches Support Different Educational Applications?

This section analyzes how prompt engineering supports a wide range of educational applications across 50 reviewed studies. A total of 62 coded instances were
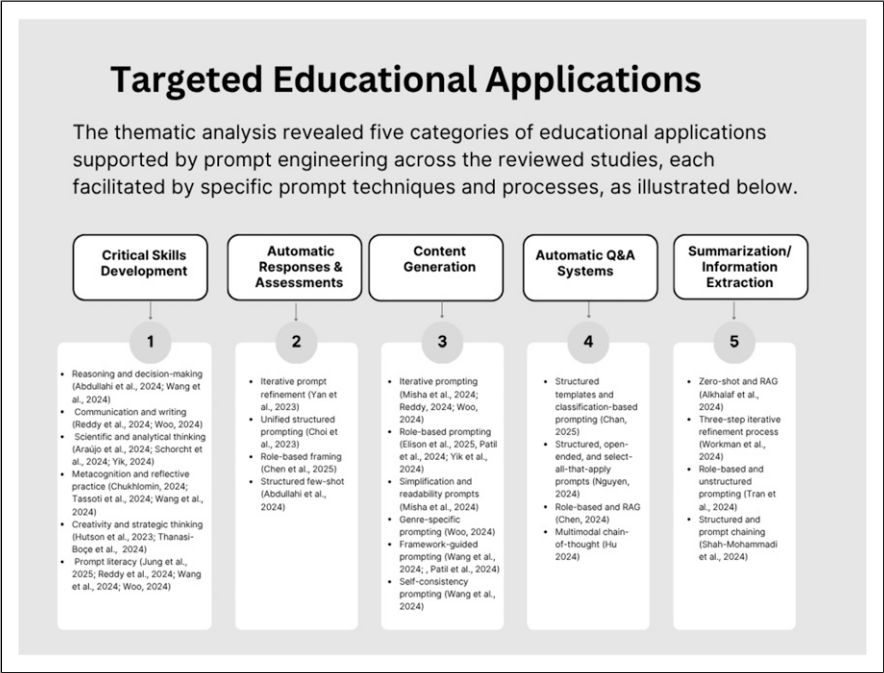
Figure 6. Educational Applications that are Supported by Prompting Approaches

identified and grouped into five overarching task categories, illustrated in Figure 6. Many studies contributed to multiple categories, reflecting the multifaceted ways in which prompt engineering facilitates learning. For instance, Haupt et al. (2024) supported both prompt literacy skill-building and automated reasoning for misinformation detection, demonstrating how a single prompt design can serve multiple educational purposes.

1. Critical Skills Development. Among the 50 studies reviewed, critical skills development emerged as the most frequently addressed learning objective, with 26 studies focused on this area. Across this body of work, generative AI was not merely used for content generation but positioned as a cognitive partner—a tool that supports iterative thinking, structured argumentation, reflective analysis, and creative exploration. While the specific critical skills varied across disciplines—from STEM to writing, art, and entrepreneurship—the unifying element was the strategic use of prompts as pedagogical tools to support deeper learning. The critical skills targeted by these studies can be organized into six key domains.

Reasoning and Decision-Making. Many studies emphasized diagnostic reasoning, evaluative judgment, and strategic decision-making. For example, Abdullahi et al. (2024) and Wang, Bi et al. (2024) had students compare AI-generated diagnoses with expert responses, rank options, and justify conclusions—developing clinical discernment through clarification cycles. Similarly, Thanasi-Boçe and Hoxha (2024) used

structured business simulations (COMIN and PREP) to help learners assess feasibility, weigh trade-offs, and refine ideas under uncertainty.

Communication and Writing. Reddy et al. (2024), Woo et al. (2024), and Jang et al. (2024) supported students in improving their writing through prompt-based collaboration with ChatGPT. Through iterative prompt-response-revision cycles, learners developed scientific writing, argument construction, and genre-specific expression, improving clarity, coherence, and contextual appropriateness.

Scientific and Analytical Thinking. Several STEM-focused studies used prompts to enhance problem-solving and analytical reasoning. Araújo and Saúde (2024) guided chemistry students through experimental design, while Schorcht et al. (2024) used Chain-of-Thought (CoT) prompting to scaffold algebraic reasoning. Yik and Dood (2024) had students model organic reaction mechanisms, fostering critical analysis of scientific concepts.

Metacognition and Reflective Practice. Metacognitive skills, such as self-monitoring and strategic revision, were emphasized in studies like Chukhlomin (2024), who used Socratic prompting for guided self-inquiry. Jung et al. (2025) and Tassoti (2024) used frameworks like CLARITY and CRISPE to help students evaluate prompt quality and refine model outputs.

Creativity and Strategic Thinking. Creative tasks were supported by prompting strategies that encouraged exploration and iteration. Hutson and Cotroneo (2023) documented how art and design students refined prompts to influence AI-generated visuals, building visual literacy and design critique skills. Thanasi-Boçe and Hoxha (2024) challenged business students to innovate and assess risk through AI-augmented simulations.

Prompt Literacy. Perhaps the most cross-cutting competency, prompt literacy refers to students' ability to craft, revise, and evaluate prompts for effective human-AI interaction. Studies by Cain (2024), Desseauve et al. (2024), Jung et al. (2025), and Woo et al. (2024) taught students how phrasing affects AI behavior and how to revise prompts to improve depth, clarity, or alignment with disciplinary goals. This literacy is increasingly recognized as a foundational skill in AI-enhanced education.

These studies employed a variety of prompting strategies, often embedded in custom instructional frameworks. Common techniques included: (1) Iterative prompt refinement (e.g., Reddy et al., 2024; Vidhani & Mariappan, 2024; Woo et al., 2024): Students cycled through prompt-and-revise loops to improve clarity and relevance. (2) Chain-of-Thought (CoT) (e.g., Lee, Teo et al., 2024; Schorcht et al., 2024): Used to scaffold step-by-step problem-solving and make reasoning visible. (3) Role-based prompting (e.g., Haupt et al., 2024; Yik, 2024): Asked learners or the AI to take on specific personas (e.g., teacher, scientist) to frame responses appropriately. (4) Scenario-based prompting (e.g., Araújo et al., 2024; Nguyen et al., 2023): Embedded prompts within real-life or simulated contexts to enhance relevance and transfer. (5) Framing and meta-prompting (e.g., Jung et al., 2025; Nguyen et al., 2023): Guided students to reflect on how prompt framing shapes the nature and tone of AI output. (6) Framework-guided prompting: CRISPE – Emphasizing Capacity, Role, Insight, Statement, Personality, and

Experiment (Tassoti, 2024), COMIN / PREP – Supporting structured reasoning and revision in entrepreneurial decision-making (Thanasi-Boçe & Hoxha, 2024), and CLARITY – Providing scaffolds for clear, targeted AI interaction in clinical reasoning (Wang et al., 2024a, 2024b).

Together, these 26 studies show that prompt engineering is not just a technical skill, it is a pedagogical strategy that redefines how learners engage with content, tools, and ideas. By treating generative AI as a learning partner, students were empowered to interrogate, co-construct, and reflect upon knowledge. Most importantly, the emphasis on prompt literacy across these studies signals its emergence as a core competency—one that educators should explicitly teach to help students ask better questions, adapt their strategies, and think critically with intelligent systems.

2. Automatic Responses and Assessments. Eleven studies in this category investigate how generative AI can produce automated responses or assessments within real-time educational and clinical settings. These systems simulate human evaluators, assistants, or communicators, with prompt engineering serving as a key mechanism for aligning AI outputs with the demands of specific contexts. Carefully crafted prompts were central to ensuring that the AI's behavior was accurate, context-aware, and pedagogically or clinically appropriate.

In clinical communication, Yan et al. (2024) demonstrated that iterative prompt refinement significantly enhanced the quality of AI-generated responses to patient medical advice requests (PMARs). By incorporating clinician and patient feedback across three revision cycles, prompt designs improved clarity, tone, and completeness. As a result, clinician acceptance of AI-generated messages rose from 62% to 84%, and patient satisfaction increased as well. Key strategies included unifying category-specific prompts into a single structured format and establishing human-in-the-loop feedback mechanisms to guide ongoing prompt optimization. This approach not only improved communication quality but also reduced clinician workload.

Similarly, Chen, Reddy et al. (2024) evaluated ChatGPT's responses to ophthalmic case studies and found the model capable of producing plausible, coherent, and safe diagnostic suggestions. Prompts were modeled on real clinical vignettes and followed standardized formats, enabling consistent outputs that aligned with professional expectations. This structured prompting approach supported formative assessment and ensured both depth and reliability in responses.

In educational contexts, Stadler et al. (2024) used generative AI to produce multiple-choice questions (MCQs) for medical training. Prompts were designed to embed clear diagnostic reasoning paths, ensuring that the generated questions matched instructional goals. Likewise, Choi et al. (2024) applied structured prompting in clinical assessment scenarios, refining prompt phrasing to elicit more accurate and consistent AI interpretations of patient data, such as in delirium evaluation tasks.

Across these studies, several prompting strategies stood out: (1) iterative refinement based on stakeholder feedback (Yan et al., 2024), (2) unified prompt structuring to replace fragmented or inconsistent systems (He et al., 2024), (3) standardized clinical

scenarios to ensure comparability and consistency (Chen et al., 2025), and (4) role-based prompting and few-shot examples to establish tone, format, and domain specificity (Haupt et al., 2024; Yik, 2024).

Collectively, these findings demonstrate that when prompt engineering is embedded in structured, feedback-informed workflows, it can significantly improve the quality, reliability, and trustworthiness of AI-generated assessments. In both education and healthcare, these systems hold strong potential to reduce workload, enhance feedback cycles, and deliver context-sensitive support at scale.

3. Content Generation. The nine studies reviewed under the "Content Generation" category focus on using generative AI to create, co-create, or support the production of domain-specific written content. These studies span disciplines such as healthcare, language education, feedback writing, and biochemistry instruction. Their emphasis lies not just in the AI's ability to generate content, but in how learners interact with AI outputs and develop prompt literacy to guide and refine the generative process.

Creating Educational and Informational Texts. In the study by Ellison et al. (2025), the researchers explored the *de novo* generation of patient education materials for colorectal conditions. The study used expert-curated prompts to instruct ChatGPT to generate texts that would be readable, accurate, and tailored to patients' comprehension levels. Prompting involved carefully calibrated role-based prompts such as "Act as a healthcare communicator" and instructional constraints (e.g., reading level set to eighth grade). Clinicians then evaluated outputs for accuracy and clarity. This study showcased how prompt specificity in tone, target audience, and formatting is essential to generating usable and trustworthy content.

Similarly, Mishra et al. (2024) examined how different prompts influenced the complexity and readability of cardiovascular disease information generated by ChatGPT. Researchers tested simplification prompts (e.g., "Explain in simple terms") and meta-prompts that explicitly requested shorter sentences and more common vocabulary. This approach highlighted how adjusting the prompt structure could lead to meaningful reductions in lexical complexity, better aligning content with the patient's understanding needs.

Supporting Student Writing through AI. Reddy et al. (2024) investigated AI-assisted writing assignments in biochemistry education. Students engaged in machine-in-the-loop writing, using scaffolded prompts and feedback cycles to refine AI-generated drafts. The process incorporated contextualized prompts (e.g., specifying target audience, tone, or citation requirements) and iterative revisions, fostering content co-creation while developing scientific writing fluency.

Woo et al. (2024) focused on EFL students composing essays with ChatGPT. The workshop included prompt engineering instruction where students learned to craft prompts suited for genre-based writing tasks (e.g., feature article or editorial letter). Prompting was framed as a genre activity: students learned to ask for outlines, vocabulary suggestions, and feedback, and to distinguish between their own contributions and the chatbot's. This genre-informed prompting supported both content generation and metacognitive writing development.

Wan and Chen (2024) examined how LLMs could aid in feedback generation. Students used prompts to request feedback on their own or peers' drafts, asking for clarity, structure, grammar, and tone evaluation. Prompts like "Review this paragraph for logical coherence" or "Give suggestions for improving the conclusion" enabled students to treat ChatGPT as a dialogic writing partner. The study highlighted feedback-focused prompting as a pedagogical approach to enhance both the quality and the self-efficacy of student revision processes.

Improving Technical Precision and Domain Alignment. Wang, Bi et al. (2024) demonstrated how prompt engineering impacts the standardization of diagnostic terminology in obstetric EMRs. The study tested four prompt designs—zero-shot, in-context, chain-of-thought (CoT), and self-consistency—to compare performance in aligning free-text diagnoses to standardized terms. Among these, self-consistency prompting (which aggregates multiple sampled outputs using majority voting) yielded the highest F1-scores. This study illustrates how structured prompt strategies significantly improve AI performance in technical text generation and classification.

Patil et al. (2024) proposed a framework-based prompting model for healthcare content generation. Their "LLM4Health" initiative was designed to prompt templates with domain-specific constraints such as clinical tone, citation formatting, and summarization scope. Prompt engineering was conceptualized as a standard operating procedure, where users selected templates based on task type—explanatory, instructional, or evaluative—thus formalizing prompt structures for consistency and content accuracy.

Across these studies, several key prompting strategies supported effective content generation: (1) Iterative prompting: Used to refine outputs over multiple prompt-response cycles (Mishra et al., 2024; Reddy et al., 2024; Shi et al., 2023; Woo et al., 2024). (2) Role-based prompting: Directed the AI to take on specific personas or communicative roles (Ellison et al., 2025; Patil et al., 2024; Yik, 2024). (3) Simplification & readability prompts: Explicitly instructed the AI to reduce complexity (Mishra et al., 2024). (4) Genre-specific prompting: Guided writing in specific forms or rhetorical structures (Woo et al., 2024). (5) Framework-guided prompting: Formal templates ensured clarity and standardization (Patil et al., 2024; Wang, Wang et al., 2024). (6) Self-consistency prompting: Aggregated outputs to ensure reliable results (Wang, Wang et al., 2024).

Content generation studies show that well-designed prompts can transform generative AI from a passive tool into a productive co-creator. The effectiveness of AI-generated outputs was closely tied to the sophistication of prompting strategies. Whether simplifying medical language, supporting EFL learners, or aiding scientific writing, these studies illustrate that *prompting is not just a technical input, but a pedagogical act* that demands awareness, iteration, and contextual adaptation.

4. Automatic Q&A Systems. The category of Automatic Q&A Systems explores how large language models can autonomously generate and respond to educational and medical queries with domain relevance, contextual understanding, and interactive

precision. These systems are used to simulate tutor-student dialogues, answer clinical questions, and produce multiple-choice or open-ended test items.

Six studies focused on building such systems across diverse domains. In education, Chan et al. (2025) developed automatic item generation for various STEM subjects using structured templates and classification-based prompting. The study emphasized zero-shot and few-shot strategies, which allowed the model to generate a range of question types across disciplines like mathematics and biology. Similarly, Nguyen et al. (2023) compared ChatGPT and Google Bard on their ability to answer academic questions across disciplines using structured, open-ended, and select-all-that-apply prompts. They found that prompt clarity and alignment with assessment goals were critical to the performance of both systems.

In healthcare, Chen, Reddy et al. (2024) introduced EyeGPT to support patient queries in ophthalmology. Techniques such as role-based prompting (e.g., "You are an ophthalmologist") and retrieval-augmented generation (RAG) were employed to embed domain-specific knowledge and ensure accurate, empathetic responses. Similarly, He et al. (2024) evaluated ChatGPT's ability to provide high-quality responses to medical inquiries and highlighted the importance of prompt specificity and user intention alignment. Hu et al. (2024) extended Q&A applications into multimodal contexts, exploring visual question answering (VQA) for medical images. Their system combined diagnostic image descriptions with targeted prompts, allowing the AI to generate interpretable outputs and support clinical reasoning across both text and visual data.

In sum, the studies reviewed demonstrate that effective Q&A performance relies on a combination of strategic prompting techniques. Few-shot prompting with annotated examples helps guide model behavior, while role-based prompting enables the simulation of expert personas. Multi-turn interactions allow for more in-depth, responsive dialogue, and structured templates ensure clarity, consistency, and alignment with domain expectations. Together, these strategies reveal that generative AI, when supported by well-crafted prompts, can move beyond static information delivery to become an adaptive, interactive tool for reasoning, learning, and decision-making in complex educational and clinical environments.

5. Summarization and Information Extraction. Although less frequently addressed, summarization and information extraction emerged as important applications of generative AI in the four reviewed studies. These tasks were explored across healthcare, journalism, and academic research, with a consistent focus on how well-crafted prompts enhance output accuracy, relevance, and structure.

In healthcare, Alkhalaf et al. (2024) investigated summarizing clinical documentation using the Llama 2 model. They employed zero-shot prompting—asking the model to summarize without prior examples—and strengthened factual grounding through retrieval-augmented generation (RAG). Carefully structured prompt templates provided domain-specific context, helping reduce hallucinations and align outputs with clinical expectations. This approach demonstrated a scalable method for generating reliable, evidence-based summaries.

Workman et al. (2024) applied a three-step iterative prompting strategy to summarize heart failure cases. The first prompt produced a broad overview, the second emphasized key clinical details (e.g., symptoms), and the third refined the summary for brevity and clarity. This staged method helped improve alignment with expert-written summaries and showed how prompt refinement can enhance both content precision and readability.

In journalism, Tran et al. (2024) explored how prompting could replicate professional newswriting. Using BERTopic to extract key topics from articles, researchers built prompts that included both keywords and role-based framing (e.g., "act as a journalist"). This strategy enabled ChatGPT to generate articles that closely mirrored human-authored news in tone, voice, and organization, allowing for direct comparison and evaluation of stylistic fidelity.

Shah-Mohammadi and Finkelstein (2024) focused on clinical information extraction through named entity recognition. They used highly structured prompts to guide the model in identifying and categorizing substance use within clinical notes. To handle this complex, multi-step task, they applied prompt chaining, where the output of one prompt (e.g., detecting alcohol use) served as the input for a follow-up classification prompt. This method provided a practical framework for navigating sensitive and layered data extraction challenges.

Across all four studies, a common conclusion emerged: the quality of summarization and extraction hinges as much on prompt design as on model capability. When tailored to context and task, prompts effectively guide large language models to generate outputs that are accurate, relevant, and aligned with domain-specific needs.

Implications for Educational Practices

This study identifies two complementary dimensions of prompt engineering—technique-based and process-based —each offering practical strategies to support the meaningful and responsible integration of AI into teaching and learning. The following section presents two summary tables, each outlining one dimension along with specific approaches and sample prompts. The sample prompts are intentionally simplified to illustrate the essential keywords and structural features characteristic of each prompt type.

Technique-Based Prompting: Designing for Function and Form. The first framework (see Table 1) categorizes prompts by how they function structurally and interactively. These techniques serve distinct pedagogical goals.

Structured Prompting. Structured prompting uses predefined patterns—such as examples, constraints, and evaluation frameworks—to guide AI responses. For example, *few-shot prompts* help AI understand the structure of a good response by offering multiple exemplars ("Here are three examples. Use them to solve the next task."). Teachers may also deploy *rubric-based prompts* for automatic grading or feedback aligned with assessment standards.

Table 1. A Guide to Technique-Based Prompting

Technique-based	Description	Features	Techniques	Prompting examples	Pedagogical uses
Structured prompting	Guide AI behavior with predefined structures	Guide model behavior using example-based scaffolds	One-shot	"Here is one example. Now try a similar one"	Support learning transfer with worked examples; fine-tune model with minimal input
			Few-shot	"Here are three examples. Use them to solve the next task"	Deepen understanding by comparing and generalizing from multiple examples
		Narrow output with format, logical, or scenario constraints	Format/structure	"List your answer in bullet points"	Help students organize thinking and present responses clearly under structure
			Conditional constraints	"If X is true, then what follows?"	Support procedural thinking in logic or scientific reasoning
			Scenario constraints	"Given a $500 budget, create a plan"	Automate assessment and align outputs to evaluation standards
		Structure output through evaluation frameworks	Multiple-choice	"Choose the best diagnosis from the list"	Automate assessment and align outputs to evaluation standards
			Ranking	"Rank these ideas from most to least effective"	Encourage evaluative thinking and prioritization
			Rubric/Metric-based	"Use the rubric to rate this response on clarity and completeness"	Provide consistent, criteria-based feedback on student outputs

(continued)

Table 1. (continued)

Technique-based	Description	Features	Techniques	Prompting examples	Pedagogical uses
Role-based/persona prompting	Instruct AI to assume specific identity or role	Support learner engagement through pedagogical identities	Educational roles	"Act as a mentor helping with a project"	Foster reflection, personalized learning, and dialogic engagement in education
		Simulate expert reasoning in clinical/professional domains	Expert roles	"You are an ophthalmologist. Explain this case"	Train domain-specific reasoning; simulate case analysis and expert explanation
		Explore influence of social identity perspectives	Social identity roles	"As a conservative user, evaluate this headline"	Examine cognitive bias and ethical implications of identity framing
		Simplify professional knowledge for lay audiences	Health educator role	"Explain this lab result in patient-friendly terms"	Improve communication of health content; foster health literacy
Contextual priming & linguistic clarity	Enhances relevance and readability through context and clear language	Build domain- or task-specific context to guide output	Contextual priming	"Based on the case details, what is the next step?"	Tailor outputs to instructional context; simulate real-world conditions
		Combine prompt with external knowledge sources	RAG	"Use this database to answer the question"	Enhance factual accuracy and depth; scaffold domain-specific reasoning
		Use context- and/or audience-specific language for clarity	Linguistic clarity	"Explain in a way a middle schooler would understand"	Support language learners and improve accessibility of explanations

Table 1. (continued)

Technique-based	Description	Features	Techniques	Prompting examples	Pedagogical uses
Chain-of-thought reasoning	Guides AI to break down complex problems step by step, focusing on reasoning	Encourage stepwise reasoning from prompt to output	Standard CoT	*"Let's solve this step by step"*	Enable step-by-step problem-solving in STEM or logic tasks
		Add structural guidance before reasoning begins	Modified CoT	*"First define key terms, then solve"*	Teach structured thinking and ordered problem resolution
		Explore multiple solution branches before answering	Tree-of-thought, self-consistency	*"Try different approaches and pick the best"*	Reveal different approaches to open-ended problems
		Use consensus across multiple reasoning paths	Majority voting	*"Aggregate multiple solutions and decide"*	Support consensus-building and collective reasoning processes
		Model iterative questioning and response refinement	Socratic CoT	*"What might be a better way to respond?"*	Foster critical thinking and metacognition through dialogic reflection
		Chain decision points for reflective business reasoning	Multi-CoT	*"Evaluate the idea, then assess risk and market"*	Support iterative thinking, planning, and business model development

(continued)

Table 1. (continued)

Technique-based	Description	Features	Techniques	Prompting examples	Pedagogical uses
Unstructured prompting	Use no prior examples or predefined structures for quick, general content	Use no prior examples; rely on model's general capability	Zero-shot prompt	"Write a short summary of this article"	Assess general language ability or initial understanding
		Explicitly instruct AI using direct or open instructions	Direct prompt	"Translate this paragraph into simple language"	Give clear, direct instructions for general, fact-based responses
		Provide specific tasks without structured reasoning	Explicit prompt	"Give three reasons for your position"	Assign basic exercises or tasks without over-specifying process
		Leave model to infer output format and logic	Open-ended prompt	"What do you think is happening here?"	Encourage exploratory and creative thinking

Role-Based/Persona Prompting. It helps AI adopt a pedagogically or professionally relevant voice. Prompts such as "Act as a mentor helping with a project" can simulate teacher-student interaction, fostering reflection and personalization. Similarly, domain-specific personas (e.g., ophthalmologists or health educators) enhance clarity and trust in instructional or clinical contexts.

Contextual Priming and Linguistic Clarity. Such prompts ensure that AI outputs are audience-aware. A prompt like "Explain in a way a middle schooler would understand" demonstrates how language adaptation supports accessibility and inclusion. This category also includes techniques like RAG (Retrieval-Augmented Generation), which combines prompts with external sources to support factual accuracy.

Chain-of-Thought Reasoning. It scaffolds logical thinking through step-by-step generation. Variants such as Tree-of-Thought or Socratic CoT promote deeper cognitive engagement by encouraging multiple solution paths or dialogic questioning ("What might be a better way to respond?"), which can enrich student reasoning in math, science, and problem-solving activities.

Unstructured Prompting. While simpler, it remains valuable for creativity and quick checks. Prompts like "Write a short summary of this article" assess baseline understanding without complex setup and are useful in exploratory or diagnostic tasks.

Together, these techniques function as a pedagogical design toolkit, not simply to elicit better responses from AI, but to actively shape the learning experience. For example, structured prompting supports efficiency, accuracy, and learning assessment by standardizing responses; role-based prompting enables simulation of expert perspectives for authentic learning; prompts emphasizing contextual clarity aid differentiation by adjusting outputs to learner needs; chain-of-thought prompting scaffolds critical thinking through step-by-step reasoning; and unstructured prompting fosters creativity, exploration, and divergent thinking. In this way, prompt engineering becomes a strategic element of instructional design, enabling educators to guide not just what AI says, but how students engage with knowledge and cognitive processes.

Process-Based Prompting: Designing for Iteration and Intent. Complementing the structural strategies, a second framework (see Table 2) addresses how prompts are developed, refined, and aligned with goals. These process-based approaches reveal the importance of prompt engineering not as a static skill, but as a reflective practice.

Iterative Refinement. It emphasizes prompt development as a cyclical process of experimentation and adjustment, mirroring the way humans naturally learn through trial, reflection, and revision. In art and design courses, students improve their visual storytelling by drafting and refining prompts with tools like DALL-E 2, gradually enhancing the nuance and clarity of their outputs. In writing instruction, learners build argumentation skills by iteratively adjusting prompt structure and tone in response to feedback. This process not only strengthens the quality of

Table 2. A Guide to Process-Based Prompting

Process-based	Sub-category	Approaches	Prompting example	Pedagogical uses
Approaches focused on designing and refining prompts to align with learning goals, performance, and task specificity	Iterative refinement	Prompt version testing	*"Revise this prompt for better diagnostic accuracy"*	Improve prompt quality through expert-informed revision and task alignment
		Scaffolded prompt writing	*"Add a follow-up question to support your argument"*	Develop writing and critical thinking skills through Scaffolded practice
		Creative iteration using generative tools	*"Experiment with phrasing to refine the image concept"*	Foster creative thinking and technical precision in design-based learning
		User-centered refinement cycles	*"Adjust the CoT structure to support collaborative reasoning"*	Support group inquiry and adapt prompts for discourse and collective learning
	Feedback-driven design	Human-in-the-loop review	*"Compare AI responses across different persona framings"*	Assess variation in output to explore model sensitivity and objectivity
		Write–curate–verify	*"Generate, review, and rate a scenario using a rubric"*	Enhance prompt effectiveness through structured educator review cycles
		Student-led prompt critique	*"Refine your image prompt to better capture the word's meaning"*	Develop visual literacy and metacognition through iterative evaluation
	Goal-oriented and task-specific framework	Rubric-aligned scoring prompts	*"Score this response using classroom assessment criteria"*	Align AI scoring with student learning outcomes using predefined criteria
		Structured entrepreneurial frameworks	*"Generate a business plan using the COMIN framework"*	Structure decision-making in entrepreneurship education
		Persona-specific self-directed learning	*"Act as a mentor to help reflect on this learning challenge"*	Promote autonomy, reflection, and emotional engagement in self-guided learning

AI-generated content but also deepens learners' understanding of rhetorical choices and communicative intent.

Feedback-Driven Design. It centers on human-in-the-loop evaluation, where educators or peers assess AI-generated outputs to inform and guide prompt revisions. This approach leverages expert judgment to refine both the content and the prompting process. For instance, the Write–Curate–Verify framework (Bai et al., 2024) incorporates structured, rubric-based validation in computational thinking courses, enabling teachers to review and verify AI-generated scenarios before classroom use. Beyond improving alignment with learning objectives, this method cultivates prompt literacy by helping students recognize how specific revisions, whether to tone, clarity, or structure can enhance the effectiveness and reliability of AI responses.

Goal-Oriented and Task-Specific Framework. It adopts a forward-engineering approach (i.e., designing prompts intentionally based on instructional goals from the outset). For example, some educators embedded prompts into assessment systems aligned with classroom rubrics, enabling automated yet pedagogically consistent scoring. Others created AI personas—such as an "inquiry coach" or "reflective mentor"—to provide targeted learning support aligned with metacognitive and socio-emotional objectives.

These process-oriented strategies elevate prompting from a technical input to an instructional design method. They encourage educators and learners to engage in *design thinking with AI*, using feedback loops, creative iteration, and performance alignment to co-develop both prompts and learning outcomes. Note that the examples in the tables are simplified to show only keywords or the gist of each prompt type. In practice, prompts would be more complete.

Task-Aware Prompt Engineering: Matching Prompting Strategies to Pedagogical Demands. Additionally, Research Question 2 identified five categories of educational uses, ranging from critical skills development to summarization and information extraction, underscoring the versatility of prompt engineering in supporting diverse educational goals. Rather than functioning as one-size-fits-all tools, prompts can be tailored to the demands of specific cognitive, procedural, and evaluative tasks. For example, structured and role-based prompts were frequently used to scaffold reasoning and analytical thinking, while self-consistency and genre-specific strategies supported content generation. In contrast, zero-shot and template-based prompting were more prevalent in summarization and information retrieval, where precision and brevity are key. This alignment suggests that educators should not only consider what content they want AI to generate but also the *nature of the learning task*—whether it's cultivating metacognition, facilitating argumentation, or assessing understanding—and select or adapt prompting strategies accordingly. Embedding task-aware prompt engineering into instructional design can help ensure that AI tools function not just as assistants but as pedagogically coherent extensions of the learning environment.

Future Directions

The landscape of prompt engineering is continuously evolving, especially with the integration of multimodal generative AI technologies. Multimodal prompt learning significantly expands traditional prompt engineering by incorporating diverse input forms such as text, images, and structured data (Liu, 2023; Liu & Chilton, 2022). This advancement directly impacts existing prompt engineering techniques and processes, as it necessitates the development of new prompting strategies capable of handling complex interactions across multiple modalities. Enhanced multimodal capabilities promise to enrich context-awareness and enable more sophisticated AI interactions, leading to improvements in tasks such as visual reasoning, cross-modal comprehension, and contextually nuanced response generation (Khattak et al., 2023). Consequently, prompt engineering techniques and iterative processes will need to be adapted to effectively leverage multimodal inputs, which may lead to the refinement and updating of the techniques and processes discussed in this study.

Another promising development is the rapid advancement of AI models with enhanced reasoning capabilities, such as GPT-4.5 and its successors, which indicates another critical avenue for prompt engineering. With these models becoming increasingly proficient in complex reasoning tasks, prompt engineering will likely shift toward leveraging advanced reasoning strategies, such as iterative chain-of-thought, self-consistency checks, and reflective dialogue prompts. These enhanced reasoning capabilities enable AI models to produce more reliable, transparent, and interpretable outputs, thereby expanding their utility in domains requiring sophisticated analytical and critical thinking skills. As AI reasoning progresses, prompt engineering will play a pivotal role in guiding these models to perform intricate, domain-specific cognitive tasks, thereby shaping a new era of interactive and intelligent human-AI collaboration.

Limitations

This systematic review has several limitations. First, the majority of studies analyzed (43 out of 50) focus on higher education contexts, with only a small fraction (7 out of 50) investigating prompt engineering within K-12 educational settings. This imbalance limits the generalizability and applicability of the findings to the broader educational spectrum, especially within primary and secondary education.

Second, preprint studies were intentionally excluded from this analysis. Although this approach ensures that all included studies underwent peer review, it potentially omitted cutting-edge and innovative findings available in preprints, which might have provided additional insights or newer methodological advances relevant to prompt engineering.

Lastly, this review primarily aimed to provide a broad overview of prompt engineering techniques, processes, and their associated educational tasks. While offering valuable insights, this approach did not specifically address the detailed

impact of these prompting methods on learning outcomes or critical skill development. As more rigorously designed empirical studies emerge, future research should focus specifically on examining how different prompt engineering strategies directly affect student learning outcomes, critical thinking, and other essential educational skills.

Conclusion

This systematic review underscores the pivotal role of prompt engineering in education by categorizing existing approaches into technique-focused and process-focused frameworks and mapping them to targeted educational uses. This classification not only clarifies how educators and learners can engage effectively with generative AI but also illustrates the versatility of prompt engineering in supporting diverse goals, including critical skills development, content generation, synthesis and extraction, automatic Q&A systems, and automated assessment. The included summary tables serve as practical tools, linking strategies to specific tasks and offering actionable examples to guide educational uses.

Prompt engineering emerges not as a technical add-on but as a core pedagogical literacy essential for meaningful AI integration. As multimodal and reasoning-rich AI systems advance, the ability to design purposeful prompts will become even more critical in shaping high-quality learning experiences. Building prompt literacy among educators and learners is, therefore, key to unlocking the transformative potential of generative AI in education.

ORCID iD

Yufeng Qian ⬤ https://orcid.org/0000-0002-7652-9467

Consent for Publication

This is a systematic literature review study.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The datasets used and analyzed during the current study are available from the author upon reasonable request.

References

Articles marked with an * are included in the systematic review.

*Abdullahi, T., Singh, R., & Eickhoff, C. (2024). Learning to make rare and complex diagnoses with generative AI assistance: Qualitative study of popular large language models. *JMIR Medical Education*, *10*(e51391), https://doi.org/10.2196/51391

*Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, *9*(2), 135–147. https://doi.org/10.1080/23735082.2023.2258886

*Alkhalaf, M., Yu, P., Yin, M., & Deng, C. (2024). Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *Journal of Biomedical Informatics*, *156*(104662), https://doi.org/10.1016/j.jbi.2024.104662

*Araújo, J. L., & Saúde, I. (2024). Can ChatGPT enhance chemistry laboratory teaching? Using prompt engineering to enable AI in generating laboratory activities. *Journal of Chemical Education*, *101*(5), 1858–1864. https://doi.org/10.1021/acs.jchemed.3c00745

*Babin, B., & Coberly, J. (2024). Prompt pattern engineering for test question mapping using ChatGPT: A cross-sectional study. *Computers and Education: Artificial Intelligence*, *5*(100165), https://doi.org/10.1016/j.caeai.2023.100165

*Bai, S., Gonda, D. E., & Hew, K. F. (2024). Write-curate-verify: A case study of leveraging generative AI for scenario writing in scenario-based learning. *IEEE Transactions on Learning Technologies*, *17*, 1301–1312. https://doi.org/10.1109/tlt.2024.3378306

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. In *NIPS'20: 34th international conference on neural information processing systems* (pp. 1877–1901). Curran Associates. https://doi.org/10.5555/3495724.3495883

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, *68*(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

*Chan, K. W., Ali, F., Park, J., Sham, K. S. B., Tan, E. Y. T., Chong, F. W. C., Qian, K., & Sze, G. K. (2025). Automatic item generation in various STEM subjects using large language model prompting. *Computers and Education: Artificial Intelligence*, *8*(100344), https://doi.org/10.1016/j.caeai.2024.100344

Chen, J. S., Reddy, A. J., Al-Sharif, E., Shoji, M. K., Kalaw, F. G. P., Eslani, M., Lang, P. Z., Arya, M., Koretz, Z. A., Bolo, K. A., Arnett, J. J., Roginiel, A. C., Do, J. L., Robbins, S. L., Camp, A. S., Scott, N. L., Rudell, J. C., Weinreb, R. N., Baxter, S. L., … Granet, D. B. (2024). Analysis of ChatGPT responses to ophthalmic cases: Can ChatGPT think like an ophthalmologist? *Ophthalmology science*, *5*(1), 100600. https://doi.org/10.1016/j.xops.2024.100600

*Chen, X., & Makmur, A. (2024). EyeGPT for patient inquiries and medical education: Enhancing AI-human communication in ophthalmology. *npj Digital Medicine*, *7*(1), 82. https://doi.org/10.1038/s41746-024-01074-z

*Chen, Y., Liu, X., & Hu, Z. (2024). Prompt engineering in K-12 STEM education: A systematic review. arXiv preprint arXiv:2410.11123. https://doi.org/10.48550/arxiv.2410.11123

*Choi, Y. K., Lin, S. Y., Fick, D. M., Shulman, R. W., Lee, S., Shrestha, P., & Santoso, K. (2024). Optimizing ChatGPT's interpretation and reporting of delirium assessment outcomes: Exploratory study. *JMIR Formative Research*, *8*(e51383), https://doi.org/10.2196/51383

*Chukhlomin, V. (2024). Socratic prompts: Engineered dialogue as a tool for AI-enhanced educational inquiry. *Labs Review*, *1*(1), 1–13. https://doi.org/10.70469/labsreview.v1i1.10

*Desseauve, D., Lescar, R., de la Fourniere, B., Ceccaldi, P.-F., & Dziadzko, M. (2024). AI in obstetrics: Evaluating residents' capabilities and interaction strategies with ChatGPT. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *302*, 238–241. https://doi.org/10.1016/j.ejogrb.2024.09.008

*Ellison, I. E., Oslock, W. M., Abdullah, A., Wood, L., Thirumalai, M., English, N., Jones, B. A., Hollis, R., Rubyan, M., & Chu, D. I. (2025). De novo generation of colorectal patient educational materials using large language models: Prompt engineering key to improved readability. *Surgery*, *180*(109024), https://doi.org/10.1016/j.surg.2024.109024

*Fernández, A., López-Torres, M., Fernández, J., & Vázquez-García, D. (2024). ChatGPT as an instructor's assistant for generating and scoring exams. *Journal of Chemical Education*, *101*(9), 3780–3788. https://doi.org/10.1021/acs.jchemed.4c00231

*Haupt, M. R., Yang, L., Purnat, T., & Mackey, T. (2024). Evaluating the influence of role-playing prompts on ChatGPT's misinformation detection accuracy: Quantitative study. *JMIR Infodemiology*, *4*(e60678), https://doi.org/10.2196/60678

*He, Z., Bhasuran, B., Jin, Q., Tian, S., Hanna, K., Shavor, C., Arguello, L. G., Murray, P., & Lu, Z. (2024). Quality of answers of generative large language models versus peer users for interpreting laboratory test results for lay patients: Evaluation study. *Journal of Medical Internet Research, 26*(e56655), https://doi.org/10.2196/56655

*Ho, B., Mayberry, T., Nguyen, K. L., Dhulipala, M., & Pallipuram, V. K. (2024). ChatReview: A ChatGPT-enabled natural language processing framework to study domain-specific user reviews. *Machine Learning With Applications*, *15*, 1–15. https://doi.org/10.1016/j.mlwa.2023.100522

*Hu, X., Gu, L., Kobayashi, K., Liu, L., Zhang, M., Harada, T., Summers, R. M., & Zhu, Y. (2024). Interpretable medical image visual question answering via multi-modal relationship graph learning. *Medical Image Analysis*, *97*(103279), https://doi.org/10.1016/j.media.2024.103279

*Hutson, J., & Cotroneo, P. (2023). Generative AI tools in art education: Exploring prompt engineering and iterative processes for enhanced creativity. *Metaverse*, *4*(1), 1–14. https://doi.org/10.54517/m.v4i1.2164

*Hwang, Y., Lee, J. H., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of new literacy skill using generative AI. ArXiv. https://arxiv.org/abs/2311.05373

*Jang, J., Eun, S., Lee, H., Choi, J., & Cho, Y. H. (2024). The effects of prompt scaffolding on learning to write arguments with ChatGPT. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), Proceedings of the 18th international conference of the learning sciences - ICLS 2024 (pp. 1502–1505). International Society of the Learning Sciences. https://doi.org/10.22318/icls2024.831011

*Jung, H., Oh, J., Stephenson, K. A. J., Joe, A. W., & Mammo, Z. N. (2025). Prompt engineering with ChatGPT 3.5 and GPT-4 to improve patient education on retinal diseases. *Canadian Journal of Ophthalmology, 60*(3), e375–e381. https://doi.org/10.1016/j.jcjo.2024.08.010

Khattak, M. U., Rasheed, H., Maaz, M., Khan, S., & Khan, F. S. (2023). MaPLe: Multi-modal prompt learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) (pp. 19113–19122). IEEE Computer Society. https://doi.org/10.1109/CVPR52729.2023.01832

*Lee, A., & Palmer, M. (2025). Prompt engineering in higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *73*(2), 245–267. https://doi.org/10.1186/s41239-025-00503-7

Lee, A. V. Y., Teo, C. L., & Tan, S. C. (2024). Prompt engineering for knowledge creation: Using chain-of-thought to support students' improvable ideas. *Artificial Intelligence*, *5*(3), 1446–1461. https://doi.org/10.3390/ai5030069

*Lee, E. Y., Dae il, N. G., An, G.-H., Lee, S., & Lim, K. (2023). ChatGPT-based debate game application utilizing prompt engineering. In Proceedings of the 2023 international conference on research in adaptive and convergent systems (pp. 1–6). ACM. https://doi.org/10.1145/3599957.3606244

*Lee, G. G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, *6*(100213), https://doi.org/10.1016/j.caeai.2024.100213

*Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, *29*(9), 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

*Lee, U., Kim, Y., Lee, S., Park, J., Mun, J., Lee, E., Kim, H., Lim, C., & Yoo, Y. J. (2024). Can we use GPT-4 as a mathematics evaluator in education? Exploring the efficacy and limitation of LLM-based automatic assessment system for open-ended mathematics question. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-024-00448-4

Liu, J., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In Proceedings of the 2022 CHI conference on human factors in computing systems (Vol. *384*, pp. 1–23). ACM. https://doi.org/10.48550/arXiv.2109.06977

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586. https://doi.org/10.48550/arXiv.2107.13586

Liu, V. (2023). Beyond text-to-image: Multimodal prompts to explore generative AI. In Extended abstracts of the 2023 CHI conference on human factors in computing systems (CHI EA'23). ACM. https://doi.org/10.1145/3544549.3577043

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

*Mishra, V., Sarraju, A., Kalwani, N. M., & Dexter, J. P. (2024). Evaluation of prompts to simplify cardiovascular disease information generated using a large language model: Cross-

sectional study. *Journal of Medical Internet Research*, *26*(e55388), https://doi.org/10.2196/55388

*Nguyen, D., MacKenzie, A., & Kim, Y. H. (2024). Encouragement vs. liability: How prompt engineering influences ChatGPT-4's radiology exam performance. *Clinical Imaging*, *115*(110276), https://doi.org/10.1016/j.clinimag.2024.110276

*Nguyen, D., Swanson, D., Newbury, A., & Kim, Y. H. (2023). Evaluation of ChatGPT and google bard using prompt engineering in cancer screening algorithms. *Academic Radiology*, *31*(5), 1799–1804. https://doi.org/10.1016/j.acra.2023.11.002

*Nguyen, P. A. (2024). Evaluating AI-generated language as models for strategic competence in English language teaching. *IAFOR Journal of Education: Studies in Education, 12*(3), 325–349. https://doi.org/10.22492/ije.12.3.13

OpenAI. (2022). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., 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. *PLoS Medicine*, *18*(3), Article e1003583. https://doi.org/10.1371/journal.pmed.1003583

Park, J., & Choo, S. (2024). Generative AI prompt engineering for educators: Practical strategies. *Journal of Special Education Technology*, *40*(3), 01626434241298954. https://doi.org/10.1177/01626434241298954

*Patil, R., Heston, T. F., & Bhuse, V. (2024). Prompt engineering in healthcare. *Electronics*, *13*(15), 2961. https://doi.org/10.3390/electronics13152961

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI Blog. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

*Reddy, M. R., Walter, N. G., & Sevryugina, Y. V. (2024). Implementation and evaluation of a ChatGPT-assisted special topics writing assignment in biochemistry. *Journal of Chemical Education*, *101*(7), 2740–2748. https://doi.org/10.1021/acs.jchemed.4c00226

Sahoo, P., Singh, A. K., Saha, S., & Jain, V. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. arXiv. https://arxiv.org/abs/2402.07927

Schick, T., & Schütze, H. (2022). True few-shot learning with prompts—A real-world perspective. *Transactions of the Association for Computational Linguistics*, *10*, 716–731. https://doi.org/10.1162/tacl_a_00485

*Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem: Investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*, *9*(1386075), https://doi.org/10.3389/feduc.2024.1386075

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, *57*(4), 620–631. https://doi.org/10.1111/ejed.12532

*Shah-Mohammadi, F., & Finkelstein, J. (2024). Extraction of substance use information from clinical notes: Generative pretrained transformer-based investigation. *JMIR Medical Informatics*, *12*(e56243), https://doi.org/10.2196/56243

*Shi, Y., Ren, P., Wang, J., Han, B., ValizadehAslani, T., Agbavor, F., Zhang, Y., Hu, M., Zhao, L., & Liang, H. (2023). Leveraging GPT-4 for food effect summarization to enhance product-specific guidance development via iterative prompting. *Journal of Biomedical Informatics*, *148*(104533), https://doi.org/10.1016/j.jbi.2023.104533

*Sivarajkumar, S., Kelley, M., Samolyk-Mazzanti, A., Visweswaran, S., & Wang, Y. (2024). An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: Algorithm development and validation study. *JMIR Medical Informatics*, *12*(e55318), https://doi.org/10.2196/55318

*Stadler, M., Horrer, A., & Fischer, M. R. (2024). Crafting medical MCQs with generative AI: A how-to guide on leveraging ChatGPT. *GMS Journal of Medical Education*, *41*(2), Doc20. https://doi.org/10.3205/zma001675

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Sage.

*Tassoti, S. (2024). Assessment of students use of generative artificial intelligence: Prompting strategies and prompt engineering in chemistry education. *Journal of Chemical Education*, *101*(6), 2475–2482. https://doi.org/10.1021/acs.jchemed.4c00212

*Thanasi-Boçe, M., & Hoxha, J. (2024). From ideas to ventures: Building entrepreneurship knowledge with LLM, prompt engineering, and conversational agents. *Education and Information Technologies*, *29*, 24309–24365. https://doi.org/10.1007/s10639-024-12775-z

*Tran, V. H., Sebastian, Y., Karim, A., & Azam, S. (2024). Distinguishing human journalists from artificial storytellers through stylistic fingerprints. *Computers*, *13*(12), 328. https://doi.org/10.3390/computers13120328

*Venerito, V., Lalwani, D., Del Vescovo, S., Iannone, F., & Gupta, L. (2024). Prompt engineering: The next big skill in rheumatology research. *International Journal of Rheumatic Diseases*, *27*(5), Article e15157. https://doi.org/10.1111/1756-185X.15157

*Vidhani, D. V., & Mariappan, M. (2024). Optimizing human–AI collaboration in chemistry: A case study on enhancing generative AI responses through prompt engineering. *Chemistry*, *6*(4), 723–737. https://doi.org/10.3390/chemistry6040043

*Wan, T., & Chen, Z. (2023). Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning. *Physical Review Physics Education Research*, *20*(1), 010152. https://doi.org/10.1103/PhysRevPhysEducRes.20.010152

*Wang, L., Bi, W., Zhao, S., Ma, Y., Lv, L., Meng, C., Fu, J., & Lv, H. (2024). Investigating the impact of prompt engineering on the performance of large language models for standardizing obstetric diagnosis text: Comparative study. *Journal of Medical Internet Research*, *26*(e12345), https://formative.jmir.org/2024/1/e53216

*Wang, M., Wang, M., Xu, X., Yang, L., Cai, D., & Yin, M. (2024). Unleashing ChatGPT's power: A case study on optimizing information retrieval in flipped classrooms via prompt engineering. *IEEE Transactions on Learning Technologies*, *17*(3), 629–640. https://doi.org/10.1109/TLT.2023.332471

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837. https://doi.org/10.48550/arXiv.2201.11903

*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*(18), 24963–24990. https://doi.org/10.1007/s10639-024-12819-4

*Workman, T. E., Ahmed, A., Sheriff, H. M., Raman, V. K., Zhang, S., Shao, Y., Faselis, C., Fonarow, G. C., & Zeng-Treitler, Q. (2024). ChatGPT-4 extraction of heart failure symptoms and signs from electronic health records. *Progress in Cardiovascular Diseases*, *87*, 44–49. https://doi.org/10.1016/j.pcad.2024.10.010

*Yan, S., Knapp, W., Leong, A., Kadkhodazadeh, S., Das, S., Jones, V. G., Clark, R., Grattendick, D., Chen, K., Hladik, L., Fagan, L., & Chan, A. (2024). Prompt engineering on leveraging large language models in generating response to InBasket messages. *Journal of the American Medical Informatics Association: JAMIA*, *31*(10), 2263–2270. https://doi.org/10.1093/jamia/ocae172

*Yik, B. J., & Dood, A. J. (2024). ChatGPT convincingly explains organic chemistry reaction mechanisms slightly inaccurately with high levels of explanation sophistication. *Journal of Chemical Education*, *101*(5), 1836–1846. https://doi.org/10.1021/acs.jchemed.4c00235

Author Biography

Yufeng Qian is an Associate Professor of Learning Analytics and Educational Technology in the College of Human Sciences and Education at Louisiana State University, United States. Her research interests include emerging technologies and pedagogy, technology leadership, and the integration of artificial intelligence in education.