Revista da Sociedade Brasileira de Ensino de Química

ISSN: 2676-0290

https://doi.org/10.56117/resbenq.2026.v7.e072609

Prompt Engineering and the critical use of Artificial Intelligence in Chemistry Education: a narrative literature review

Engenharia de Prompt *e o uso crítico de Inteligência Artificial no Ensino de Química: uma revisão narrativa da literatura*

Ingeniería de Prompts *y el uso crítico de la Inteligencia Artificial en la Enseñanza de la Química: una revisión narrativa de la literatura*

Guilherme Gonçalves Costa (*guilherme.g.costa@hotmail.com*)
Instituto de Química (IQ) – Universidade Estadual de Campinas (UNICAMP)
https://orcid.org/0000-0002-8530-6847

Feliphe Cristiano Belluci (*feliphe.cristiano@gmail.com*)
Instituto de Química (IQ) – Universidade Estadual de Campinas (UNICAMP)
https://orcid.org/0009-0009-7038-8614

Gildo Girotto Júnior (*ggirotto@unicamp.br*)
Instituto de Química (IQ) – Universidade Estadual de Campinas (UNICAMP)
https://orcid.org/0000-0001-9933-100X

Abstract

The rapid spread of Generative Artificial Intelligence (GAI) in educational contexts has raised expectations and tensions, yet there is still limited systematization on how to guide this interaction through Prompt Engineering (PE) strategies. Accordingly, this article aimed to identify and analyze, through a Narrative Literature Review (NLR), the main conceptions, approaches, and/or PE strategies associated with the use of GAI in Chemistry Education, discussing their foundations, potential, and limits. To this end, we searched four databases (Web of Science, Scopus, ERIC, and Google Scholar), followed by duplicate removal, application of exclusion criteria, and full-text reading of the remaining records. The final *corpus* comprised seven studies, all international, reinforcing both the emergent character of the topic and the limited body of publications aligned with this scope. To identify and classify PE strategies, we adopted the taxonomy proposed by Chen *et al.* (2025) as a theoretical-methodological basis and conducted an analytical and interpretive reading of the articles. The results indicate a predominance of basic PE strategies, suggesting that PE has been used more often as a pragmatic condition for employing GAI than as a deliberate pedagogical object. The only advanced strategy identified across the studies was the strategy named Chain-of-Thought. Still, two specific strategies—"Five S" Prompting and Prompt Cueing—signal efforts toward educational recontextualization. In the discussion, we organize the findings into guiding propositions, including a proposed definition of PE as a didactic-discursive artifact that translates teachers' choices regarding objectives and quality criteria, as well as a digital competence to be taught. Finally, we systematize prompt examples in Portuguese to support situated decisions by teachers and students, from request design to verification and formative feedback.

Keywords: Pedagogical strategies. Artificial Intelligence. Digital technologies. Chemistry Education.

Resumo

A rápida disseminação das Inteligências Artificiais Generativas (IAG) em contextos educacionais tem suscitado expectativas e tensões, mas ainda há pouca sistematização sobre como orientar essa interação por meio de estratégias de Engenharia de *Prompt* (EP). Assim, este artigo objetivou identificar e analisar, por meio de uma Revisão Narrativa da Literatura (RNL), as principais concepções, abordagens e/ou estratégias de EP associadas ao uso das IAG no Ensino de Química, discutindo seus fundamentos, potencialidades e limites. Para isso, realizamos buscas em quatro bases (*Web of Science*, Scopus, ERIC e Google Acadêmico), seguidas de remoção de duplicatas, aplicação de critérios de exclusão e leitura integral dos registros remanescentes. O *corpus* final foi composto por sete estudos, todos internacionais, o que reforça o caráter emergente do tema e a baixa produção alinhada a esse escopo. Para identificar e classificar estratégias de EP, adotamos a taxonomia de Chen *et al.* (2025) como base teórico-metodológica e conduzimos uma leitura analítica e interpretativa dos artigos. Os resultados apontam para a predominância de estratégias básicas de EP, sugerindo que esta tem sido mais mobilizada como condição pragmática de uso das IAG do que como objeto pedagógico deliberado. A única estratégia avançada identificada entre os trabalhos foi a estratégia denominada *Chain-of-Thought*. Ainda assim, duas estratégias específicas, *"Five S" Prompting* e *Prompt Cueing*, sinalizam esforços de recontextualização educacional. Na discussão, organizamos os achados em proposições orientadoras, entre as quais sugerimos uma definição de EP como artefato didático-discursivo que traduz escolhas docentes sobre objetivos e critérios de qualidade, bem como como competência digital a ser ensinada. Por fim, sistematizamos exemplos de *prompts* em português, visando apoiar decisões situadas de docentes e estudantes desde o desenho da solicitação até a verificação e a devolutiva formativa.

Palavras-chave: Estratégias pedagógicas. Inteligência Artificial. Tecnologias Digitais. Ensino de Química.

Resumen

La rápida difusión de las Inteligencias Artificiales Generativas (IAG) en contextos educativos ha suscitado expectativas y tensiones, pero aún existe poca sistematización sobre cómo orientar esta interacción mediante estrategias de Ingeniería de *Prompt* (IP). Así, este artículo tuvo como objetivo identificar y analizar, mediante una Revisión Narrativa de la Literatura (RNL), las principales concepciones, enfoques y/o estrategias de IP asociadas al uso de IAG en la Enseñanza de la Química, discutiendo sus fundamentos, potencialidades y límites. Para ello, realizamos búsquedas en cuatro bases (Web of Science, Scopus, ERIC y Google Académico), seguidas de la eliminación de duplicados, la aplicación de criterios de exclusión y la lectura integral de los registros restantes. El *corpus* final estuvo compuesto por siete estudios, todos internacionales, lo que refuerza el carácter emergente del tema y la baja producción alineada con este alcance. Para identificar y clasificar estrategias de IP, adoptamos la taxonomía de Chen *et al.* (2025) como base teórico-metodológica y realizamos una lectura analítica e interpretativa de los artículos. Los resultados señalan el predominio de estrategias básicas de IP, lo que sugiere que la IP ha sido movilizada más como condición pragmática de uso de las IAG que como objeto pedagógico deliberado. La única estrategia avanzada identificada entre los trabajos fue la estrategia denominada *Chain-of-Thought*. Aun así, dos estrategias específicas, *"Five S" Prompting* y *Prompt Cueing*, señalan esfuerzos de recontextualización educativa. En la discusión, organizamos los hallazgos en proposiciones orientadoras, entre las cuales sugerimos una definición de la IP como artefacto didáctico-discursivo que traduce elecciones docentes sobre objetivos y criterios de calidad, así como una competencia digital que debe enseñarse. Por último, sistematizamos ejemplos de *prompts* en portugués, con el fin de apoyar decisiones situadas de docentes y estudiantes desde el diseño de la solicitud hasta la verificación y la devolución formativa.

Palabras clave: Estrategias Pedagógicas. Inteligencia Artificial. Tecnologías Digitales. Enseñanza de la Química.

Introduction

Nowadays, we are living in an era in which technological resources enable us to carry out multiple tasks simultaneously: we reply to personal messages, handle work emails, and continuously check social media. In this context, Santaella (2024) associates the ubiquity of technologies—that is, our constant connectivity to today's Web—with its trend of presentism. Such an excessive focus on the "now" is further driven by the volume and speed of information generated through human and algorithmic production in cyberspace.

It is within this large volume of data—or, more commonly, big data—that Web 4.0 or even 5.0 (Santaella, 2024) encompasses prominent themes of recent years, such as cloud computing, the Internet of Things, the Internet of Bodies, and smart cities. All these themes are further intensified by discussions around Artificial Intelligence (AI) and, in particular, Generative AI (hereafter, GenAI). The rapid popularization of GenAI has been reconfiguring practices of textual production, information seeking, and the authorship of texts, images, and videos (Chen *et al.*, 2025).

From this perspective, GenAI-based chatbots operate from the user's input of an initial text (known as a "prompt"), returning solutions, content, and data through large pretrained models. From late 2022 onward, with the spread of conversational systems such as ChatGPT (Helder & Oliveira, 2022), the presence of these tools has also expanded in school and university contexts (Nicolau, 2025), whether through institutional initiatives (Nunes, 2024) or the spontaneous adoption by teachers and students (Alves *et al.*, 2025).

Accordingly, GenAI began to be mobilized for diverse purposes, such as synthesizing and explaining content, creating exercises, lesson planning, supporting academic writing, tutoring, and feedback (Alier *et al.*, 2024; Gonçalves Costa *et al.*, 2024). In parallel, equally relevant concerns have been consolidated, such as persuasive yet incorrect answers (Gupta *et al.*, 2023), biases and stereotypes (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022), opacity of operation (Sampaio *et al.*, 2024), cognitive dependence (Cardoso *et al.*, 2023), privacy violations (Sampaio *et al.*, 2024), authorship conflicts (Rodrigues & Rodrigues, 2023), and uncritical uses that weaken pedagogical intentionality (Leite, 2023).

We consider it pertinent to emphasize that interaction with GenAI is not reduced to a query logic typical of search engines (such as Google or Bing), in which the user receives and selects previously indexed results. In generative systems, the input prompt participates in the very production of the response, so that small variations in how a request is phrased can substantially change the content, the level of detail, and the reliability of what is generated. For this reason, the way of "asking" becomes part of the pedagogical problem. It can mitigate or amplify the risks already mentioned, such as persuasive inaccuracies, biases, opacity, and uncritical uses. It is within this horizon that the discussion of Prompt Engineering (PE) as an emerging practice is situated.

Prompt Engineering (PE) as an emerging practice

Operationally, a prompt is the natural-language input command through which the user formulates a request so that GenAI produces an output (Sampaio *et al.*, 2024). In GenAI systems based on Large Language Models (LLMs), this production results from a probabilistic process of token prediction (minimal units of text), in which the initial prompt reconfigures the space of possibilities and therefore influences the type of textual continuation generated (Chen *et al.*, 2025; Lima-Lopes, 2025). In this sense, Prompt Engineering (PE) refers to adjusting a prompt to make it more appropriate, enabling a language model to perform a given task (Liu *et al.*, 2023). We further note that any improvement in the output does not, in itself, guarantee conceptual validity, transparency of reasoning, or pedagogical appropriateness.

That said, in the educational field, it is advisable to avoid a technicist reading of PE as a "command" or "trick" for optimization. In teaching contexts, prompts also function as didactic-discursive artifacts, as they translate choices about learning objectives, conceptual scope, quality criteria, expected language, and ways of monitoring students' work. From this standpoint, it is useful to treat PE less as an optimization procedure (particularly the typically prescriptive approaches—very common in the current debate) and more as a repertoire of strategies that can be mobilized didactically and pedagogically.

Recent literature has pointed to a plurality of strategies associated with PE, ranging from more elementary procedures to more complex approaches (Chen *et al.*, 2025). Liu *et al.* (2023), for example, identify five basic considerations for effective PE:
(1) pretrained language model choice; (2) prompt template engineering; (3) prompt answer engineering; (4) expanding the paradigm; and (5) prompt-based training strategies. Federiakin *et al.* (2024), from an educational perspective, understand the PE process as a multidimensional skill consisting of: (1) understanding prompt structure; (2) prompt literacy; (3) prompt design methods; and (4) online critical reasoning with LLMs.

In this article, we take as a reference the taxonomy of PE strategies mapped in the review conducted by Chen *et al.* (2025), largely because of its synthesizing and comprehensive character. This choice aims to ensure relative alignment with a body of work that is still very recent and rapidly expanding, while not losing sight of the need for reinterpretations in light of the educational field (Federiakin *et al.*, 2024), especially when the goal is to discuss didactic-pedagogical implications, as is the case here. To make this repertoire visible, Table 1 summarizes the main PE strategies based on the taxonomy systematized by Chen *et al.* (2023).

Table 1

Main Prompt Engineering (PE) strategies

PE Strategies	Description
Providing Instructions (PI)	Drafting guidelines that steer a model's behavior and output.
Being Clear and Precise (BCP)	Using specific, unambiguous instructions in the prompt to reduce ambiguity and guide more relevant/correct responses.
Iterative Refinement (IR)	Manually and continuously adjusting the prompt based on a critical evaluation of earlier GenAI outputs.
Role-Based Prompting (RBP)	Assigning the model a role (e.g., teacher, reviewer, tutor) to guide style, focus, and response criteria.
Use of Delimiters (Del)	Using markers (e.g., """...""", <...>, [[...]]) to separate instructions from data/content and reduce ambiguity.
Trying Several Times (TST)	Running the same prompt multiple times to generate alternatives and then selecting/combining the best based on criteria.
Zero-Shot Prompting (ZSP)	Issuing a prompt with no examples, relying only on the model's prior knowledge to perform the task.
One-Shot Prompting (OSP)	Including a single input–output example to guide format and expected type of result.
Few-Shot Prompting (FSP)	Including multiple examples to provide context and guide performance, especially for complex/ambiguous tasks.
Chain-of-Thought (CoT)	Requesting/structuring step-by-step reasoning before the final answer, supporting logic/problem-solving tasks.
Self-Consistency (SC)	Generating multiple answers with different reasoning paths and selecting the most recurrent/coherent one.
Generated Knowledge Prompting (GKP)	Asking the model to generate relevant background/context first, then using it to produce a more coherent final answer.
Least-to-Most Prompting (LtM)	Decomposing a complex task into simpler subproblems solved sequentially, using each step to support the next.
Tree of Thoughts (ToT)	Exploring multiple reasoning branches before answering, evaluating alternatives to pick the best solution.
Graph of Thoughts (GoT)	Organizing candidate ideas/solutions as nodes connected by dependencies, enabling navigation/combination across paths.
Decomposed Prompting (DECOMP)	Modular decomposition into subtasks handled by specialized "handlers," coordinated by a controller that integrates outputs.
Active Prompting (ActP)	Strategically selecting examples, having humans annotate them, and reusing them as one-/few-shot examples to improve later inferences.

PE Strategies	Description
Prompt Pattern Catalogs (PPC)	Documented collections of reusable prompt templates/patterns to adapt and combine for different tasks, reducing trial-and-error.
Prompt Optimization (POpt)	Often automated methods that iteratively refine prompts based on performance to improve accuracy, relevance, and alignment.
Retrieval Augmentation or Retrieval Augmented Generation (RAG)	Enriching prompts with externally retrieved information (docs/databases/web) to ground answers and reduce hallucinations.
Reasoning and Active Interaction (RAI)	Multi-step reasoning plus (when needed) tool/actions to improve accuracy, logical coherence, and contextual adequacy.

Source: Chen *et al.* (2025).

Why does Chemistry Education require responsible prompts?

In the context of GenAI, the production of highly sophisticated textual responses coexists with conceptual inaccuracies and weak inferences (Leite, 2023). This underscores the need for care in how one requests, delimits, and validates an answer produced by this type of tool. In Chemistry, this care is particularly critical, because school chemistry knowledge articulates models and idealizations (Ferreira & Justi, 2008), requires terminological precision, operates across different levels of representation (Wartha & Rezende, 2016), and depends on validity conditions that are often omitted in generic explanations (Talanquer, 2026).

Moreover, in Chemistry, learning situations frequently require students to monitor and justify inferential chains—for example, when following the logic of a stoichiometric calculation, interpreting the dynamic nature of chemical equilibrium, or sustaining a mechanistic line of reasoning for a reaction. Thus, the quality of interaction with GenAI is not exhausted by the "good formulation" of the request; it also involves criteria for verification and mediation. Along these lines, we argue that prompt construction has a didactic-pedagogical status for GenAI users, because it (i) organizes the intended task; (ii) makes users' expectations explicit; and (iii) encourages disciplinary checking.

At the same time, it is important to stress that PE, although necessary, is not sufficient in itself. It should be understood as a set of knowledges or competencies that must be articulated with other forms of teachers' professional knowledge, especially conceptual mastery of Chemistry and well-defined pedagogical criteria. Only then can GenAI function as a didactic-pedagogical tool rather than an uncritical shortcut, preserving the teacher's fundamental role in the ethical and formative guidance of these technologies.

For this to happen, however, professional education for the critical use of Digital Information and Communication Technologies (DICT) is foundational. This movement has already been taking shape through the proposal of professional development initiatives led by some Brazilian universities (Fundação de Apoio à Pesquisa do Distrito Federal [FAPDF], 2025; Universidade Federal de Minas Gerais [UFMG], 2025; Universidade Tecnológica Federal do Paraná [UTFPR], n.d.).

Contribution of the present article

This article seeks to fill a systematization gap by organizing a set of PE strategies for Chemistry Education, linking them to didactic-pedagogical purposes, limitations, and validation precautions, and proposing directions that support situated decision-making by students and teachers. Its contribution, of a theoretical-practical nature, is to offer guiding propositions and prompt examples that sustain teacher mediation and foster conscious uses of GenAI in Chemistry Education, in dialogue with contemporary challenges related to teacher education and the governance of AI use in education (UNESCO, 2022).

Research Questions (RQs) and Objective

Considering this framing, we set out to investigate the following Research Questions (RQs):

RQ1. What conceptions, approaches, and/or strategies of Prompt Engineering (PE) have been discussed in the Brazilian and international literature on the use of Generative Artificial Intelligence (GenAI) in Chemistry Education?

RQ2. How can these conceptions, approaches, and/or strategies be reinterpreted and exemplified in the Brazilian context so as to support conscious and reflective pedagogical practices?

Accordingly, the objective of this article was to identify and analyze, through a Narrative Literature Review, the main conceptions, approaches, and/or strategies of Prompt Engineering (PE) associated with the use of GenAI in Chemistry Education, discussing and systematizing their foundations, potentialities, and limits, with a view to supporting a critical, pedagogical, and contextualized use of these tools.

Although we recognize that there are distinctions among the lexical choices of the terms "conceptions," "approaches," and "strategies" (Alves & Bego, 2020), from this point onward we will use the term "strategy" in a broad sense, encompassing the others.

Methodology

This study falls within the scope of exploratory qualitative research (Gil, 2008) and consists of a Narrative Literature Review (NLR) (Merga & Oddone, 2025).

In NLRs, there is greater methodological flexibility for conducting a literature review when compared to Systematic Literature Reviews. There are also standardized protocols for their conduct aimed at reducing biases within the research team, making this a methodological approach characterized by its reproducibility (Merga & Oddone, 2025).

Moreover, they are particularly valuable for secondary research that requires clarification and proposition-building, in which a more interpretive and discursive synthesis of the existing literature is needed. They also challenge positivist and post-positivist perspectives of science that produce a single reality through experimental research by contributing a more interpretive, critical, and dynamic approach (Merga & Oddone, 2025).

In this NLR, aiming to answer the RQs, we conducted searches in Google Scholar, Web of Science, Scopus, and the Education Resources Information Center (ERIC) on January 5, 2026. We adopted a temporal scope from November 2022 to December 2025, considering the release and rapid popularization of ChatGPT starting in November 2022 (Helder & Oliveira, 2022). Since then, an intensification of debates and investigations has been observed regarding the uses, potentialities, and limits of ChatGPT and other GenAI tools in different fields, including education (Gonçalves Costa *et al.*, 2024; Hwang & Chang, 2023; Lee & Palmer, 2025). That said, Table 2 presents the search strategies used in the databases selected for this study.

Table 2

Indexing databases consulted, languages, and search strings

Indexing databases	Languages	Search strings
Google Scholar	Portuguese	(“engenharia de prompt” OR “engenharia de prompts”) AND (“ensino de química” OR “educação química”)
Web of Science Scopus ERIC	English	“prompt engineering” AND (“chemistry teaching” OR “chemistry education” OR “chemical education”)

Source: Authors.

Due to the large number of works retrieved in the initial searches (~300 studies), we opted for a search restricted to Titles, Abstracts, and Keywords. In the case of Google Scholar, we employed the search string in Portuguese to increase sensitivity to potentially national production, since Web of Science and Scopus did not return records using the Portuguese string and ERIC retrieved only one article in Portuguese, which was not relevant to Chemistry Education. Google Scholar also does not allow filtering by Titles, Abstracts, and Keywords.

By doing so, we reduced the number of records retrieved to 23 and, after extracting the works from the databases, we adopted the incorporation of all retrieved records as the sole inclusion criterion. We chose to use this broader criterion for two reasons: (i) in terms of workload, the analysis of 23 records was feasible for the research team; and (ii) we were interested in examining all records retrieved from the national literature. We then applied the following exclusion criteria: (E1) the study does not address Prompt Engineering as a research problem; (E2) the study is not related to Chemistry Education; and (E3) the study provides limited depth with respect to the RQs. Finally, we conducted full-text readings of the remaining works and systematized the information in order to answer the RQs and propose directions for Chemistry Education.

For the stage of analysis and categorization of the findings, we adopted the taxonomy proposed by Chen *et al.* (2025) as our main theoretical-methodological framework. Our choice is grounded in the comprehensive and synthesizing nature of that study, which presents a collection of PE strategies described in recent literature, ranging from their foundations to the presentation of more complex techniques.

The use of this framework enabled us to classify the PE strategies identified in the Chemistry Education literature into three analytical categories: *Basic*, *Advanced*, and *Specific*. Regarding the nomenclature of the first two categories, this was a grouping used by the authors of the study themselves (Chen *et al.*, 2025). The last category, in turn, was defined based on the specificities of two strategies that do not fit unequivocally into those indicated by Chen *et al.* (2025). Further details are provided in the Results and Discussion section.

Finally, the entire analytical process of coding the retrieved records, identifying PE strategies, and subsequently classifying them was carried out using a spreadsheet containing the complete systematization of the material accessed, including authorship, year of publication, title, journal or source, indexing database, among other information. Both the spreadsheet and the analytical process *per se* underwent peer validation, in which a researcher experienced in the topic and methodological approach assessed the appropriateness of the analytical procedures. The first two authors of this text independently analyzed the retrieved records and, in cases of disagreement, discussion was held until consensus was reached.

Results and Discussion

The search procedures resulted in 23 records across the four databases consulted, distributed as follows: Web of Science (n = 4), Scopus (n = 5), ERIC (n = 3), and Google Scholar (n = 11). After removing duplicates, 18 works remained: Web of Science (n = 4), Scopus (n = 1), ERIC (n = 2), and Google Scholar (n = 11). Next, after applying the exclusion criteria and conducting full-text readings of the remaining studies, 7 works were selected to compose the final *corpus* of this NLR, drawn from Web of Science (n = 4), Scopus (n = 1), and ERIC (n = 2), with no studies selected from Google Scholar (n = 0). Table 3 below provides the complete list of works after duplicate removal, along with their respective codes, sources, and inclusion/exclusion indications.

Table 3

Records retrieved from the consulted databases after duplicate removal

Code	Authorship	Title	Journal or source	Included?	Exclusion criteria
A1	Araújo & Saúde (2024)	Can ChatGPT Enhance Chemistry Laboratory Teaching? Using Prompt Engineering to Enable AI in Generating Laboratory Activities	Journal of Chemical Education (ISSN 1938-1328)	Yes	—
A2	Araújo *et al.* (2024)	Uso do ChatGPT no planeamento do Ensino de Química – Potencialidades e desafios da IA para a Educação em Ciências e Tecnologia	APEduC Revista/ APEduC Journal (ISSN 2184-7436)	No	E3
A3	Eugenio *et al.* (2023)	Inteligência artificial frente à resolução de exercícios de Química: um estudo exploratório com o ChatGPT	Colloquium Humanarum (ISSN 1809-8207)	No	E1
A4	Feldman-Maggor *et al.* (2025)	Perspectives of Generative AI in Chemistry Education Within the TPACK Framework	Journal of Science Education and Technology (ISSN 1573-1839)	Yes	—
A5	Reddy *et al.* (2024)	Implementation and Evaluation of a ChatGPT-Assisted Special Topics Writing Assignment in Biochemistry	Journal of Chemical Education (ISSN 1938-1328)	Yes	—
A6	Tassoti (2024)	Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education	Journal of Chemical Education (ISSN 1938-1328)	Yes	—
A7	Vasconcelos & Frota (2025)	IA como Aliada da Criatividade Humana: O Desenvolvimento do Aplicativo "Terra e Universo" com Auxílio do ChatGPT	Revista Brasileira de Informática na Educação (ISSN 2317-6121)	No	E1
A8	Vidhani & Mariappan (2024)	Optimizing Human–AI Collaboration in Chemistry: A Case Study on Enhancing Generative AI Responses through Prompt Engineering	Chemistry (ISSN: 2624-8549)	Yes	—
A9	White *et al.* (2023)	Assessment of chemistry knowledge in large language models that generate code	Digital Discovery (ISSN: 2635-098X)	Yes	—

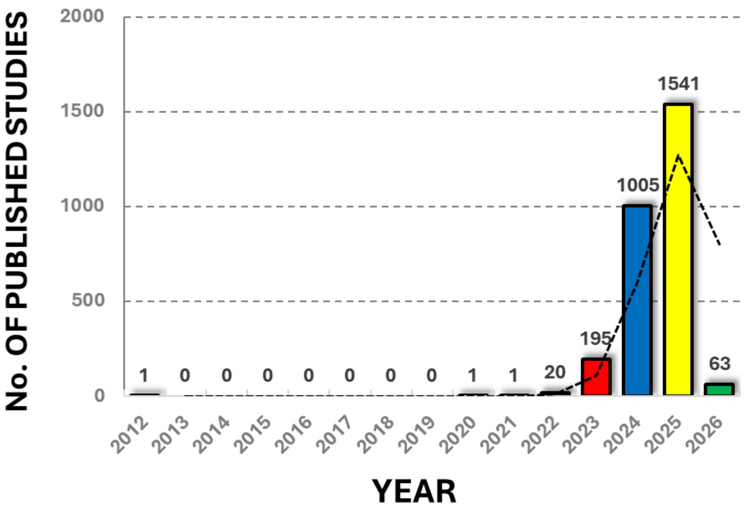
Code	Authorship	Title	Journal or source	Included?	Exclusion criteria
A10	Yik & Dood (2024)	ChatGPT Convincingly Explains Organic Chemistry Reaction Mechanisms Slightly Inaccurately with High Levels of Explanation Sophistication	Journal of Chemical Education (ISSN 1938-1328)	Yes	—
Cf1	Santos (2023)	O processo de transposição didática pelo ChatGPT acerca dos modelos atômicos: inteligência artificial como professor do Ensino Médio	Proceedings of XII Encontro Paulista de Pesquisa em Ensino de Química	No	E1
Ch1	Silva & Santos (2025)	Ensino religioso e educação de surdos: falas, práticas e diversidades na sala de aula	Book chapter	No	E1
M1	Almeida (2025)	A cibercultura no ensino de química: empregando o chat GPT para desenvolver uma sequência didática sobre o aquecimento global	Monograph	No	E1
M2	Carvalho (2024)	Utilização da inteligência artificial para o Ensino de Ciências da Natureza: uma revisão bibliográfica	Monograph	No	E1
M3	Costa (2025)	Integrando a inteligência artificial generativa na Educação em Química: desenvolvimento de ferramentas e avaliação como recurso educacional	Master's Dissertation	No	E1
M4	Luna (2025)	Integrando o ensino socrático com LLMs: um *framework* multiagente para inovação no ensino de programação	Master's Dissertation	No	E2
M5	Santos (2025)	Concepções de professores e estudantes do ensino médio sobre inteligência artificial na educação	Master's Dissertation	No	E1
M6	Silva (2024)	Estudo exploratório e análise comparativa de ferramentas de inteligência artificial generativa para o ensino de computação	Monograph	No	E2

Note: A = Scientific article; Cp = Conference proceeding; Ch = Book chapter; M = Monograph, dissertation, or thesis. Source: Authors.

As can be inferred from Table 3, all 11 works from the Brazilian literature identified were excluded from the analysis (3 articles, 2 conference proceedings, 1 book chapter, 3 monographs, and 3 master's dissertations). The fact that no Brazilian work met our criteria to provide answers to the RQs reinforces the emerging nature of the topic, since PE itself is a field of inquiry that is still consolidating. As shown in Figure 1, a search for the term "prompt engineering" in the Web of Science database reveals only three works published before 2022, the year ChatGPT was released (Helder & Oliveira, 2022). The graph in Figure 1 was built considering works published up to January 9, 2026, the date on which this text was written.

Figure 1

Trend in the number of works indexed in the Web of Science database containing the term "prompt engineering" in Titles, Abstracts, or Keywords



Source: Authors.

Notably, PE is a topic of growing interest to the academic community. Given its recent nature, we already expected that any connections to Chemistry Education would still be incipient. Thus, among the 18 records obtained (Table 3), only 7 met the criteria required for the RQs to be answered with reasonable adequacy.

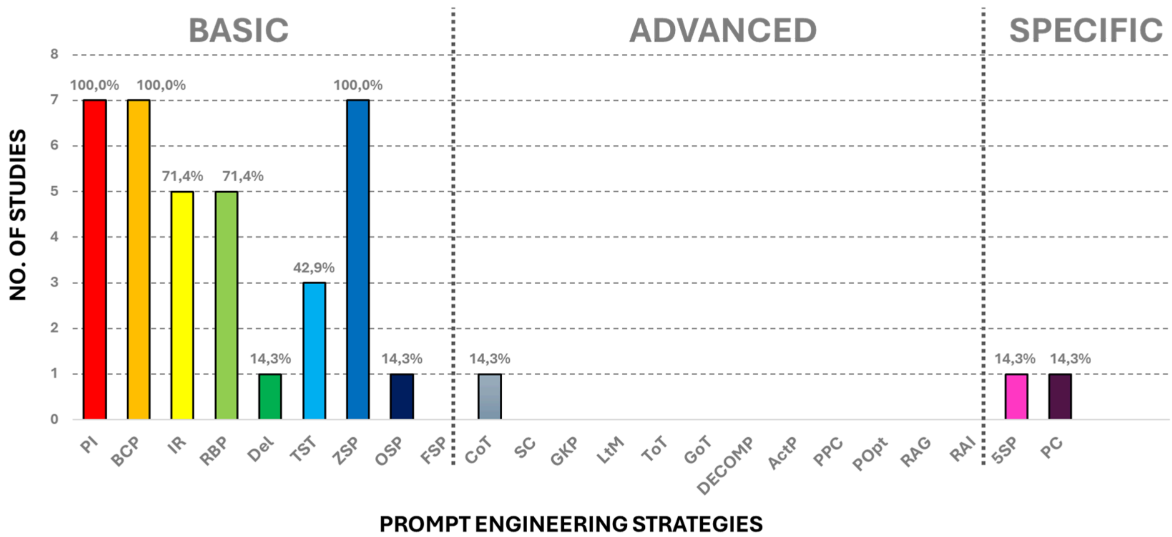
Prompt Engineering* for, in, *and* with *Chemistry Education

Drawing on the works selected for analysis, this subsection seeks to answer RQ1: "*What conceptions, approaches, and/or Prompt Engineering (PE) strategies have been discussed in the Brazilian and international literature on the use of Generative Artificial Intelligence (GenAI) in Chemistry Education?*".

At the outset, given the exclusion of all Portuguese-language works retrieved by the search, no PE strategy was explicitly discussed in the Brazilian literature in connection with Chemistry Education. With regard to the international literature, however, the seven works comprising the *corpus* met these requirements. As mentioned earlier, this study relied on the taxonomy presented in Chen *et al.* (2025) to identify and classify PE strategies (Table 1). Based on that taxonomy, Figure 2 presents a chart with the absolute number of works and their respective percentages for the identification and classification of each PE strategy across the seven studies included in the analysis.

Figure 2

Absolute numbers and relative percentages of Basic, Advanced, *and* Specific *Prompt Engineering (PE) strategies identified in the seven analyzed studies*



Legend:
PI – Providing Instructions; **BCP** – Being Clear and Precise; **IR** – Iterative Refinement; **RBP** – Role-Based Prompting; **Del** – Use of Delimiters; **TST** – Trying Several Times; **ZSP** – Zero-Shot Prompting; **OSP** – One-Shot Prompting; **FSP** – Few-Shot Prompting; **CoT** – Chain-of-Thought; **SC** – Self-Consistency; **GKP** – Generated Knowledge Prompting; **LtM** – Least-to-Most; **ToT** – Tree-of-Thoughts; **GoT** – Graph-of-Thoughts; **DECOMP** – Decomposed Prompting; **ActP** – Active Prompting; **PPC** – Prompt Pattern Catalogs; **POpt** – Prompt Optimization; **RAG** – Retrieval Augmentation; **RAI** – Reasoning and Active Interaction; **5SP** – “Five S” Prompting; **PC** – Prompt Cueing.

Source: Authors.

Also in Figure 2, the strategies were grouped into *Basic*, *Advanced*, and *Specific*. To distinguish between *Basic* and *Advanced*, we used the classification suggested by Chen *et al.* (2025) themselves. The strategies considered *Specific* were those featuring particularities not mentioned in Chen *et al.* (2025).

Because most studies do not explicitly state PE strategies according to Chen *et al.*'s (2025) taxonomy, the classification was conducted through an analytical and interpretive reading of each article. On this basis, we assessed that articles A6 (Tassoti, 2024) and A10 (Yik & Dood, 2024) were the only ones to address more specific PE strategies (Figure 2) beyond those mentioned by Chen *et al.* (2025).

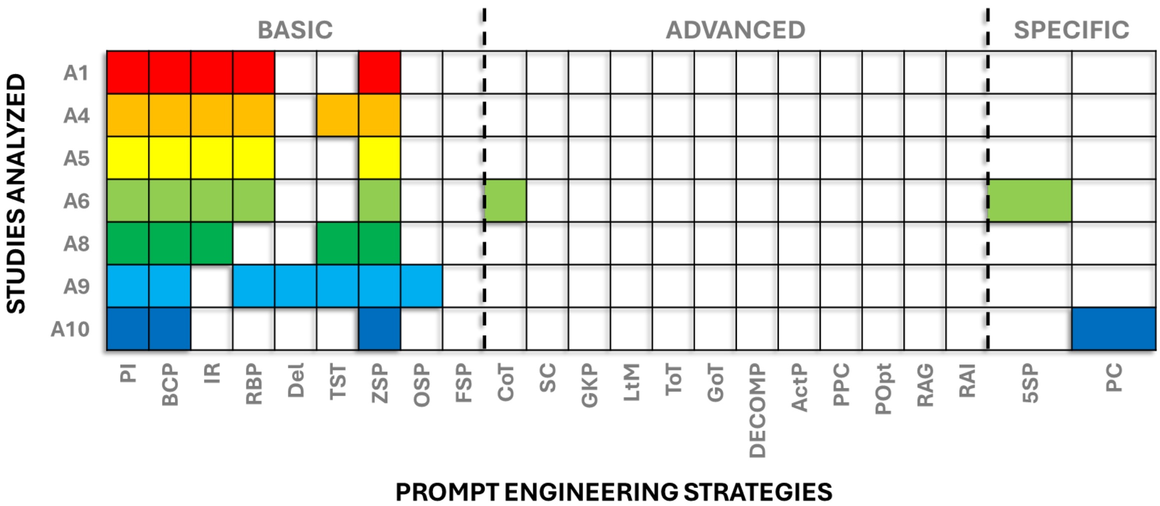
In the case of A6, Tassoti (2024) proposes the "Five S" Prompting (5SP) strategy, which combines five strategies listed in Chen *et al.* (2025), namely: (i) Providing Instructions (PI)—referred to in A6 as "Be Specific"; (ii) Being Clear and Precise (BCP)—referred to as "Simplify the Language"; (iii) Iterative Refinement (IR)—referred to as "Provide Feedback"; (iv) Role-Based Prompting (RBP)—referred to as "Set the Scene"; and (v) Chain-of-Thought Prompting (CoT)—referred to, in a simplified format, as "Structure the Output". Although these five strategies are autonomous, Tassoti's (2024) recontextualization for Chemistry Education highlights a valuable gesture of translating a repertoire formerly circumscribed to Computer Science into a situated pedagogical use.

In a similar direction, article A10 (Yik & Dood, 2024) proposes what it calls Prompt Cueing (PC). This strategy can be understood as a refinement of the prompt through structuring, based on hierarchical "cues" embedded in the prompt itself. In this case, such cues function as gradual supports to guide the type of explanation expected, eliciting descriptions on reaction mechanisms. Moreover, the strategy is of interest to the educational field because it seeks to bring the prompt-writing process closer to the Vygotskian concept of pedagogical scaffolding (Wertsch, 1991).

To synthesize this mapping and make visible how PE strategies are distributed across the included studies, Figure 3 presents a matrix crossing each work with the respective strategies identified.

Figure 3

Matrix of Basic, Advanced, *and* Specific *Prompt Engineering (PE) strategies identified in the seven analyzed studies*



Legend:
PI – Providing Instructions; **BCP** – Being Clear and Precise; **IR** – Iterative Refinement; **RBP** – Role-Based Prompting; **Del** – Use of Delimiters; **TST** – Trying Several Times; **ZSP** – Zero-Shot Prompting; **OSP** – One-Shot Prompting; **FSP** – Few-Shot Prompting; **CoT** – Chain-of-Thought; **SC** – Self-Consistency; **GKP** – Generated Knowledge Prompting; **LtM** – Least-to-Most; **ToT** – Tree-of-Thoughts; **GoT** – Graph-of-Thoughts; **DECOMP** – Decomposed Prompting; **ActP** – Active Prompting; **PPC** – Prompt Pattern Catalogs; **POpt** – Prompt Optimization; **RAG** – Retrieval Augmentation; **RAI** – Reasoning and Active Interaction; **5SP** – "Five S" Prompting; **PC** – Prompt Cueing.

Source: Authors.

As shown in Figures 2 and 3, the repertoire of PE strategies mobilized in the *corpus* tends to concentrate on procedures considered basic by Chen *et al.* (2025), especially those aimed at making explicit what is expected from the response (Providing Instructions—PI), reducing ambiguities (Being Clear and Precise—BCP), and progressively adjusting the input prompt (Iterative Refinement—IR). This pattern suggests that, in the analyzed set, PE has most often been activated as a pragmatic condition for using GenAI—that is, as a means of making responses more adequate to the task—rather than as a deliberate pedagogical object, designed in relation to learning objectives, criteria for the validity of chemical knowledge, and ways of monitoring students' reasoning.

In addition to these strategies, Role-Based Prompting (RBP) also appears recurrently, indicating that assigning specific roles and identities to GenAI has functioned as a resource for guiding the type of language, the level of detail, and the framing of the requested explanations. Taken together, PI, BCP, IR, and RBP seem to constitute pillars of interaction with GenAI in the analyzed *corpus*, as they operate mainly at the level of initial direction-setting and the progressive adjustment of generated responses.

In the analysis, the Zero-Shot Prompting (ZSP) strategy also stands out, having been identified in all studies. Its higher prevalence can possibly be explained by the fact that it is a more common form of interaction with GenAI, relying only on the model's prior knowledge and without providing examples (Chen *et al.*, 2025).

Regarding advanced strategies, their presence is quite limited. Only study A6 (Tassoti, 2024) employed—albeit in a simplified way—the Chain-of-Thought (CoT) strategy, which induces the model to make explicit a sequential reasoning chain before presenting the final answer, commonly triggered by the command "let's think step by step." Although this type of request may support tasks that require justifications and inferential chaining, its low occurrence in the *corpus* suggests that such uses have not yet consolidated in the field.

Also noteworthy is the emphasis on human oversight as an essential component of GenAI use in connection with Chemistry Education—an aspect that is relatively consensual among sources A1, A4, A5, and A8. In A4 (Feldman-Maggor *et al.*, 2025) and A8 (Vidhani & Mariappan, 2024), this oversight acts as a kind of output curation, aimed at identifying inconsistencies and correcting hallucinations, positioning PE less as automation and more as a practice of pedagogical mediation. Work A1 (Araújo & Saúde, 2024) shows an anticipated task design in which the specialist conducts multiple iterations of testing and refining the prompt before generating the final protocols (Iterative Refinement—IR). Finally, in A5 (Reddy *et al.*, 2024), oversight appears as a process, proposing the integration of students as active participants in a six- to nine-week pathway that involves "prompt clinics," peer review, and validation of evidence in the scientific literature.

From these perspectives, PE is not reduced to the pragmatic calibration of output; rather, it becomes integrated—at different levels of explicitness—into the very design and facilitation of activities. Still, when one considers the *corpus* as a whole, this incorporation appears uneven. In some works, it approaches a formative content and a teaching competence to be developed (Feldman-Maggor *et al.*, 2025), whereas in others it remains more closely tied to the curation and validation of what was generated (White *et al.*, 2023).

In summary, the findings suggest that the literature still oscillates between, on the one hand, understanding PE as a set of procedures to improve GenAI responses and, on the other, treating it as a situated pedagogical practice, articulated with validity criteria and processes for monitoring learning.

This oscillation reveals a relevant tension for Chemistry Education. Although part of the literature highlights PE as a means of making GenAI responses more useful, complete, or suited to the task, this pragmatic improvement does not necessarily amount to the pedagogical or conceptual qualification of the interaction. More structured prompts may generate textually sophisticated responses that are still inaccurate, incomplete, or insufficiently sensitive to the validity conditions of chemical knowledge. This tension is close to broader debates on GenAI in education, in which the potential of these tools to support learning and teaching work is acknowledged, while risks such as cognitive dependence, the weakening of critical thinking, biases, plausible responses without deep understanding, authorship issues, and the need for continuous human supervision are also emphasized (Farrokhnia *et al.*, 2024; Kasneci *et al.*, 2023; Tlili *et al.*, 2023). Thus, an important pedagogical implication of the findings is that PE should not be incorporated as a form to obtain better responses from GenAI, but rather as an object of formative problematization, through which students and teachers can discuss limits, omissions, authorship, safety, and responsibility for the knowledge produced. In this sense, the divergences and gaps identified in the literature reinforce the need to shift the focus from prompt efficiency to the quality of the didactic-pedagogical mediation that guides, interprets, and validates the use of GenAI in Chemistry Education.

On the basis of this reading, we propose to understand PE, within the scope of this article, through three complementary and overlapping emphases—not as rigid categories, but as ways of situating the role of prompts in relation to Chemistry Education. First, we speak of PE for Chemistry Education when prompt construction functions as a means to support tasks of learning and teaching Chemistry, designing requests with criteria and validity conditions that guide the type of response expected. Second, we address PE in Chemistry Education when the very formulation, justification, and evaluation of prompts and responses is thematized as an object of education, so as to support the validation of explanations in chemical language, considering the models, representations, and assumptions mobilized. Finally, we propose thinking of PE with Chemistry Education as a mediated and distributed collaborative practice, in which criteria and validation processes are constructed in the human–GenAI interaction with active student participation, under explicitly stated pedagogical and ethical responsibilities.

It should be emphasized that these three emphases do not assume that mastering PE is sufficient to solve the challenges of GenAI use in Chemistry Education. The didactic-pedagogical quality of interaction depends, rather, on broader professional knowledge resulting from the articulation among content knowledge, pedagogical knowledge, and technological knowledge, which supports decisions about objectives, quality criteria, limits of validity, and the monitoring of students' reasoning (Celik, 2023; Girotto Júnior *et al.*, 2025; Mishra & Koehler, 2006). In this framework, PE figures as knowledge associated with GenAI use—that is, as a component that gains meaning when subordinated to formative purposes and the concrete conditions of teaching. Thus, when considering GenAI in connection with Chemistry Education, it becomes necessary to make explicit not only "how to write prompts," but also how PE integrates with pedagogical choices, disciplinary validation, and ethical responsibility for the knowledge mobilized.

In light of this, to synthesize the findings and make explicit some implications for Chemistry Education, we systematized in Table 4 a set of potentialities, limits, and conditions for the critical use of PE when conceived for, in, and with Chemistry Education.

Table 4

Potentialities, limits, and conditions for the critical use of Prompt Engineering (PE) for*,* in*, and* with *Chemistry Education*

Dimension	
PE ***for*** Chemistry Education	**Potentialities** • Designing prompts with well-defined objectives, level, and output format to support learning and teaching tasks. • Producing explanations, exercises, and feedback with gradation and contrast. • Tailoring responses to different models, idealizations, and levels of representation. **Limits** • Generating convincing answers with conceptual gaps. • Omitting validity conditions. • Undue simplifications and excessive focus on the final product. **Conditions for critical use** • Defining what counts as an adequate answer (chemical criteria, response format, etc.). • Requiring justifications and checks (counterchecks, dimensional consistency, comparisons, etc.). • Recording decisions and reviewing use in light of observed inconsistencies.
PE ***in*** Chemistry Education	**Potentialities** • Treating prompts and responses as objects of education. • Developing GenAI literacy, argumentation, metacognition, etc. • Supporting the validation of chemical explanations (considering models, representations, etc.). **Limits** • Risk of a technicist focus and “recipes”; shifting attention toward textual performance. • Inequalities in access to and repertoire in digital information and communication technologies (ICT). • Authorship tensions. **Conditions for critical use** • Assessing the process (criteria, revision, justification, etc.) with explicit rubrics. • Working with counterexamples and comparing responses. • Planning activities with varying levels of difficulty and transparent forms of participation.
PE ***with*** Chemistry Education	**Potentialities** • Mediated and distributed collaboration between people and GenAI. • Co-constructing criteria for hypotheses, comparing explanations, revision, and guided inquiry. • Engaging students in validation and argumentation. **Limits** • Dependence and outsourcing of reasoning; automation bias. • Superficial validation by plausibility; privacy and authorship risks. • Erasing the process when GenAI “solves” the task for the user. **Conditions for critical use** • Defining roles and responsibilities (GenAI should function as a resource, not an authority). • Adopting validation and documentation protocols for the entire decision-making process. • Designing tasks that include GenAI-produced answers which students must critique and validate.

Source: Authors.

Theoretical–practical contributions for a critical use of GenAI in the Brazilian context

In this subsection, we shift our focus toward a propositional reading of the findings, aimed at constructing directions for critical uses of PE conceived *for*, *in*, and *with* Chemistry Education. Thus, we seek to answer RQ2: "*How can PE conceptions, approaches, and/or strategies be reinterpreted and exemplified in the Brazilian context so as to support conscious and reflective pedagogical practices?*".

To this end, in a move similar to the NLR conducted by Gupta *et al.* (2023), we attempt to translate this reading into suggestions for action, organized as three guiding propositions. We emphasize that these propositions should not be read as a closed set of rules or "recipes" for prompts. Rather, they operate as action cues and reflection criteria to guide didactic–pedagogical choices throughout the entire GenAI use cycle—from designing the request to validation and formative feedback—under the concrete conditions of the classroom.

PE as a didactic-discursive artifact and a formative competence.

By shifting PE into the domain of teacher planning, it is important to treat it as a conscious act of the teacher (Alves & Bego, 2020), shaped by pedagogical and ethical decisions, rather than as an operational expedient for using GenAI. From this perspective, understanding it as a "technique" tends to suggest fixed and replicable procedures, which brings the debate back toward a technicist rationality. For this reason, it is more productive to think of PE as a repertoire of strategies adjustable to the educational context (Alves & Bego, 2020). In parallel, integrating technologies into teaching requires more than instrumental mastery; it demands criteria to decide, justify, and evaluate choices (Girotto Júnior *et al.*, 2025). It is from this framing that we formulate our first proposition.

Proposition 1: *In Chemistry Education, it is productive to define PE as a didactic-discursive artifact that translates the teacher's choices regarding learning objectives and quality criteria. It should be taught as a digital competence that involves formulating, revising, and critically evaluating interactions with GenAI.*

Thematizing PE in teacher education and digital literacy.

The articulation of PE with Chemistry Education cannot be treated merely as a punctual addition to teachers' technical repertoire, because it can mobilize professional knowledge that connects content, pedagogy, and technology, as systematized by the Technological Pedagogical Content Knowledge (TPACK) model (Mishra & Koehler, 2006). In recent debate, this framework has been challenged by the need to include AI-specific competencies and ethical dimensions. In this sense, Celik (2023) proposes expanding TPACK by incorporating ethical aspects associated with the use of AI tools and the evaluation of automated decisions.

In reviews on AI in education, it is also emphasized that ethical challenges and gaps in teacher education limit integration, reinforcing TPACK as a reference to guide critical and effective uses of these technologies (de Siqueira *et al.*, 2025). In the specific case of Chemistry, Feldman-Maggor *et al.* (2025) argue that content knowledge and PCK directly influence both prompt formulation and output evaluation, and that PE relates to the technological component of TPACK but additionally requires AI literacy, including attention to biases and hallucinations.

It should be clarified, however, that the mobilization of TPACK in this article is not intended to confine teacher education to a single theoretical model, nor to suggest that this is the only possible way to articulate PE, GenAI, and Chemistry Education. On the contrary, we understand TPACK as an analytical framework that is particularly relevant to the aims of this study, as it makes it possible to elucidate the interdependence among content, pedagogical, and technological knowledge involved in the formulation, evaluation, and validation of prompts in teaching contexts.

Thus, TPACK is assumed here as a productive interpretive possibility, consistent with our theoretical scope and with the frameworks mobilized by the research team, without excluding other approaches that are equally relevant for discussing teacher education, digital literacy, criticality, ethics, and professional development. With this in mind, we arrive at our second proposition.

Proposition 2: *Effective articulation of PE with Chemistry Education can be productively interpreted in light of the teachers' TPACK, which is essential for formulating strategies involving GenAI. In this context, a well-developed TPACK supports the construction of prompts that use discipline-specific jargon, contributing to the validation and precision of generated responses. In addition, awareness of AI* *biases and hallucinations should be integrated as a new dimension of teachers' digital literacy.*

Collaboration with GenAI under validation and safety criteria.

When considering collaboration with GenAI in Chemistry activities—especially when there are experimental implications—it becomes decisive to make validation, safety, and sustainability criteria explicit, rather than treating PE merely as a way to "improve" generated text. In the case of laboratory protocols, Araújo and Saúde (2024) show that ChatGPT can produce understandable proposals but with relevant scientific–pedagogical limitations, especially regarding safety and sustainability, even omitting critical information such as toxicity and flammability. In a convergent perspective, Vidhani and Mariappan (2024) argue that accuracy depends on careful communication and iterative prompt design, but emphasize that human oversight is an essential component to safeguard reliable information and sustain a productive human–GenAI partnership. Yik and Dood (2024), in turn, show that different levels of prompt structuring can yield highly sophisticated explanations for reaction mechanisms without necessarily translating into greater mechanistic accuracy, reinforcing the need for disciplinary checking. In the Brazilian context, Leite (2023) also notes that the tool should be understood as support rather than as a substitute for the teacher, requiring critical reflection on its educational uses. It is within this horizon of collaboration mediated by validation and safety criteria that we arrive at the third and final proposition.

Proposition 3: *PE articulated with Chemistry Education must be inseparable from a disciplinary verification protocol carried out by an expert (teacher), because GenAI often fails in critical aspects of safety, sustainability, and experimental logic (for example, by completing reactions and suggesting hazardous reagents). PE should encourage checking and argumentative consistency, not merely textual adjustment.*

We believe that the propositions presented can function as an interpretive axis to translate the review findings into applicable directions. Finally, to make them more tangible, we present examples of prompt formulation and rewriting that make explicit objectives, quality criteria, and validation procedures, seeking to foster situated and ethically informed didactic–pedagogical decisions in Chemistry Education. The complete examples are systematized in Appendix A.

Limitations

This study has important limitations. First, it addresses a recent and rapidly evolving topic, which resulted in a small and exclusively international *corpus*, limiting the generalizability of the findings. Second, because it is an NLR, the study does not reach the same level of rigor and reproducibility as a systematic review (for example, regarding protocol, formal appraisal of methodological quality, and risk of bias). Third, the classification of PE strategies relied on interpretive reading, since many studies do not state them according to a single taxonomy, which may introduce ambiguities. Finally, the review did not empirically assess effects on learning or assessment, indicating the need for future studies with more robust designs and investigations in Brazilian educational contexts.

Final considerations and Future perspectives

This NLR indicates that, in recent scholarship on GenAI in Chemistry Education, PE has been mobilized mainly to fit the output to the task, with a predominance of basic strategies. In a field sensitive to conceptual precision, validity conditions, safety, and sustainability, this finding reinforces that "good phrasing" cannot be equated with a valid answer.

As an interpretive synthesis, Table 4 brought together potentialities, limits, and conditions for a critical use of PE when conceived *for*, *in*, and *with* Chemistry Education. This reading unfolds into three propositions: (i) PE *for* Chemistry Education, by treating it as a didactic-discursive artifact and a teachable digital competence, shifting prompts and responses to the status of a formative object (Proposition 1); (ii) PE *in* Chemistry Education, by making explicit that its effectiveness depends on TPACK and critical digital literacy, a sine qua non condition for formulating requests aligned with content and validating the precision of what is generated (Proposition 2); and (iii) PE *with* Chemistry Education, by sustaining human–GenAI collaboration under disciplinary verification protocols, so as to encourage checking and argumentative consistency rather than merely textual adjustment (Proposition 3). As a theoretical–practical contribution, the article translates these directions into prompt examples, aiming to support situated decisions by teachers and students from request design through verification and formative feedback.

In sum, future perspectives include: (1) expanding and updating the mapping of the field with more rigorous reviews, such as systematic ones; (2) developing and testing, in the Brazilian context, short and explicit scripts for checking GenAI output in Chemistry (e.g., units, assumptions, validity conditions, and laboratory safety); (3) investigating, through empirical studies, how students learn to justify, compare, and validate GenAI responses in chemical language; and (4) proposing and evaluating teacher education initiatives that integrate PE, ethics, authorship, privacy, and biases.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

References included in the Narrative Literature Review are marked with an asterisk (*).

Alier, M., García Peñalvo, F. J., & D. Camba, J. (2024). Generative Artificial Intelligence in Education: From Deceptive to Disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*(5), 5–14. https://doi.org/10.9781/ijimai.2024.02.011

Almeida, D. P. de. (2025). *A cibercultura no ensino de química: empregando o chat GPT para desenvolver uma sequência didática sobre o aquecimento Global* [Trabalho de Conclusão de Curso, Universidade Federal do Rio de Janeiro]. https://pantheon.ufrj.br/handle/11422/27030

Alves, A. C., Viana, H., Lima, J., Manhani, C., & Netto, V. (2025, 31 de janeiro). Ensino começa a integrar inteligência artificial no Brasil; especialistas veem oportunidade, mas com riscos. *g1*. https://g1.globo.com/politica/noticia/2025/01/31/ensino-comeca-a-integrar-inteligencia-artificial-no-brasil-especialistas-veem-oportunidade-mas-com-riscos.ghtml

Alves, M., & Bego, A. M. (2020). A celeuma em torno da temática do planejamento didático-pedagógico: definição e caracterização de seus elementos constituintes. *Revista Brasileira de Pesquisa em Educação em Ciências*, 71–96. https://doi.org/10.28976/1984-2686rbpec2020u7196

*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

Araújo, J. L., Saúde, I., Costa, A. P., & Lopes, J. B. (2024). Uso do ChatGPT no planeamento do Ensino de Química – Potencialidades e desafios da IA para a Educação em Ciências e Tecnologia. *APEduC Revista/ APEduC Journal*, *5*(2), 133–152. https://doi.org/10.58152/APEduCJournal.569

Cardoso, F. S., Pereira, N. da S., Braggion, R. C., Chaves, P., & Andrioli , M. (2023). O uso da Inteligência Artificial na Educação e seus benefícios: uma revisão exploratória e bibliográfica. *Revista Ciência Em Evidência*, *4*(FC), e023002. https://doi.org/10.47734/rce.v4iFC.2332

Carvalho, C. M. B. (2024). *Utilização da inteligência artificial para o ensino de ciências da natureza : uma revisão bibliográfica* [Trabalho de Conclusão de Curso, Instituto Federal do Espírito Santo *Campus* Colatina]. https://repositorio.ifes.edu.br/handle/123456789/6405

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. *Patterns*, *6*(6), 101260. https://doi.org/10.1016/j.patter.2025.101260

Costa, D. G. M. da. (2025). *Integrando a inteligência artificial generativa na educação em química: desenvolvimento de ferramentas e avaliação como recurso educacional* [Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte]. https://repositorio.ufrn.br/items/b18e1c57-9ef0-4780-bf3e-31e01c01f4f4

de Siqueira, L. E., Bedin, E., & Grando, J. W. (2025). Inteligência Artificial na Educação: revisão cienciométrica e proposta de articulação pelo modelo TPACK. *Educação, Formação & Tecnologias*, *13*(1), 59–85. https://doi.org/10.5281/zenodo.16741209

Eugenio, I. D., Silva, L. A. N. da, Santos, C. M. dos, & Gibin, G. B. (2023). Inteligência artificial frente a resolução de exercícios de Química: um estudo exploratório com o ChatGPT. *Colloquium Humanarum, 20*, 461–476. https://doi.org/10.5747/ch.2023.v20.h571

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International,* *61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Federiakin, D., Molerov, D., Zlatkin-Troitschanskaia, O., & Maur, A. (2024). Prompt engineering as a new 21st century skill. In *Frontiers in Education* (Vol. 9, p. 1366434). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1366434

*Feldman-Maggor, Y., Blonder, R., & Alexandron, G. (2025). Perspectives of generative AI in chemistry education within the TPACK framework. *Journal of Science Education and Technology, 34*, 1–12. https://doi.org/10.1007/s10956-024-10147-3

Ferreira, P. F. M., & Justi, R. D. S. (2008). Modelagem e o "fazer ciência". *Química Nova na Escola*, *28*, 32–36. http://qnesc.sbq.org.br/online/qnesc28/08-RSA-3506.pdf

Fundação de Apoio à Pesquisa do Distrito Federal. (2025, 29 de maio). *IAgora Brasil: UnB e FAPDF unem forças para capacitar professores da rede pública no uso humanizado da inteligência artificial*. https://www.fap.df.gov.br/w/iagora-brasil-unb-e-fapdf-unem-forcas-para-capacitar-professores-da-rede-publica-no-uso-humanizado-da-inteligencia-artificial

Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.

Girotto Júnior, G., Morais, C., Fernandes Goes, L., & Gonçalves Costa, G. (2025). A proposal to integrate teacher digital competencies and professional knowledge associated with PCK and TPACK. *Góndola, Enseñanza y Aprendizaje de las Ciencias*, *20*(2). https://doi.org/10.14483/23464712.22023

Gonçalves Costa, G., Nascimento Júnior, W. J. D., Mombelli, M. N., & Girotto Júnior. (2024). *Revisiting a Teaching Sequence on the Topic of Electrolysis: A Comparative Study with the Use of Artificial Intelligence*, *101*(8), 3255–3263. https://doi.org/10.1021/acs.jchemed.4c00247

Gupta, B., Mufti, T., Sohail, S. S., & Madsen, D. Ø. (2023). ChatGPT: A brief narrative review. *Cogent Business & Management*, *10*(3). https://doi.org/10.1080/23311975.2023.2275851

Helder, D., Oliveira, L. de. (2022, 08 de dezembro). ChatGPT: conheça o robô conversador que viralizou por ter resposta para (quase) tudo. *g1*. https://g1.globo.com/tecnologia/noticia/2022/12/08/chatgpt-conheca-o-robo-conversador-que-viralizou-por-ter-resposta-para-quase-tudo.ghtml

Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

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

Leite, B. S. (2023). Inteligência Artificial e Ensino de Química: uma análise propedêutica do ChatGPT na definição de conceitos químicos. *Química Nova*, *46*(9), 915–923. https://doi.org/10.21577/0100-4042.20230059

Lima-Lopes, R. E. de. (2025). Por uma revisão crítica do uso de Inteligência Artificial na Educação. *Revista de Estudos de Cultura*, *11*(27), 39–65. https://doi.org/10.32748/revec.v11i27.22719

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, *55*(9), 1–35. https://doi.org/10.1145/3560815

Luna, G. de S. (2025). *Integrando o ensino socrático com LLMs: um framework multiagente para inovação no ensino de programação* [Dissertação de Mestrado, Universidade Federal do Rio de Janeiro]. https://www.cos.ufrj.br/index.php/en/publicacoes-pesquisa/details/20/3232

Merga, M. K., & Oddone, K. (2025). 'This review is useful': The novel knowledge mobilisation contributions of narrative reviews in higher education. *Issues in Educational Research*, *35*(2), 698–715. https://www.iier.org.au/iier35/merga.pdf

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Nicolau, A. (2025, 15 de setembro). Uso de inteligência artificial cresce no Brasil e impacta educação. *CNN Brasil*. https://www.cnnbrasil.com.br/educacao/uso-de-inteligencia-artificial-cresce-no-brasil-e-impacta-educacao/

Nunes, T. (2024, 18 de outubro). Unicamp lança plataforma de inteligência artificial. *Portal Unicamp*. https://unicamp.br/noticias/2024/10/18/unicamp-lanca-plataforma-de-inteligencia-artificial/

*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

Rodrigues, O. S., & Rodrigues, K. S.. (2023). A inteligência artificial na educação: os desafios do ChatGPT. *Texto Livre*, *16*, e45997. https://doi.org/10.1590/1983-3652.2023.45997

Sampaio, R. C., Sabbatini, M., & Limongi, R. (2024). *Diretrizes para o uso ético e responsável da inteligência artificial generativa: um guia prático para pesquisadores*. Intercom. https://prpg.unicamp.br/wp-content/uploads/sites/10/2025/01/livro-diretrizes-ia-1.pdf

Santaella, L. (2024). Diagnóstico do contemporâneo. *Estudos Avançados*, *38*(110), 7–18. https://doi.org/10.1590/s0103-4014.2024.38110.002

Santos, C. M. dos, Silva, L. A. N. da, Barboza, T. B., & Gibin, G. B. (2023). O processo de transposição didática pelo ChatGPT acerca dos modelos atômicos: inteligência artificial como professor do Ensino Médio. *Anais do XII Encontro Paulista de Pesquisa em Ensino de Química*. IFSP Sertãozinho/USP Ribeirão Preto. https://eppeq.ti.srt.ifsp.edu.br/eppeq/article/view/97

Santos, J. V. P. dos. (2025). *Concepções de professores e estudantes do ensino médio sobre inteligência artificial na educação* [Dissertação de Mestrado, Universidade Federal de Sergipe]. https://ri.ufs.br/handle/riufs/24009

Silva, G. C. de J., & Santos, O. P. (2025). Ensino religioso e educação de surdos: falas, práticas e diversidades na sala de aula. In Gonçalves, M. C. da S. & Oliveira, M. R. N. de. (Orgs). *Educação: Reflexões e Experiências Volume 13* (pp. 58–64). Editora Poisson. http://doi.org/10.36229/978-65-5866-602-8

Silva, J. E. O. (2024). *Estudo exploratório e análise comparativa de ferramentas de inteligência artificial generativa para o ensino de computação* [Trabalho de Conclusão de Curso, Universidade Federal de Uberlândia]. https://repositorio.ufu.br/handle/123456789/44921

Talanquer, V. (2026). The Importance of Fostering Conceptual Reframing in Chemistry Education. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D5RP00421G

*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

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

United Nations Educational, Scientific and Cultural Organization. (2022). *Recomendação sobre a Ética da Inteligência Artificial*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_por

Universidade Federal de Minas Gerais. (2025, 16 de julho). *UFMG oferta 800 vagas em cursos gratuitos de educação digital*. https://www.ufmg.br/comunicacao/assessoria-de-imprensa/releases/oportunidades/ufmg-oferta-800-vagas-em-cursos-gratuitos-de-educacao-digital/

Universidade Tecnológica Federal do Paraná. (n.d.). *Inteligência artificial aplicada à educação*. Recuperado em 9 de janeiro de 2026, de https://utfpr.curitiba.br/iaeduc/

Vasconcelos, V. R. C., & Frota, D. A de. (2025). IA como Aliada da Criatividade Humana: O Desenvolvimento do Aplicativo “Terra e Universo” com Auxílio do ChatGPT. *Revista Brasileira de Informática na Educação*, *33*, 451–471. https://doi.org/10.5753/rbie.2025.5204

*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

Wartha, E. J., & Rezende, D. de B. (2016). Os níveis de representação no Ensino de Química e as categorias da semiótica de Peirce. *Investigações Em Ensino De Ciências*, *16*(2), 275–290. https://ienci.if.ufrgs.br/index.php/ienci/article/view/230

Wertsch, J. V. (1991). *Voices of the mind: Sociocultural approach to mediated action*. Harvard University Press.

*White, A. D., Hocky, G. M., Gandhi, H. A., Ansari, M., Cox, S., Wellawatte, G. P., Sasmal, S., Yang, Z., Liu, K., Singh, Y., & Peña Ccoa, W. J. (2023). Assessment of chemistry knowledge in large language models that generate code. *Digital Discovery, 2*, 368–376. https://doi.org/10.1039/D2DD00087C

*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

Submetido em: 13/01/2026 **Aceito em:** 28/07/2026 **Publicado em:** 25/09/2026

Periódico organizado pela Sociedade Brasileira de Ensino de Química – SBEnQ
Sociedade Brasileira de Ensino de Química

Appendix A

Examples of naïve prompts (NP) and strategic prompts (SP), including users' objectives and the Prompt Engineering (PE) strategies mobilized, in teaching, learning, and assessment situations in Chemistry

User Objective (student)	**PE strategies mobilized**
Obtain a personalized study plan for the Chemistry exam, with explanations and exercises at an appropriate level, assuming that GenAI acts as a Chemistry tutor and provides one solved example as a model.	Role-Based Prompting (RBP) and One-Shot Prompting (OSP)
Naïve Prompt 1 (NP_1)	
Explain Chemistry for my exam and create some questions with an answer key.	
Strategic Prompt 1 (SP_1)	
You are a private high school Chemistry tutor. Your role is to help me study for an exam with a focus on conceptual understanding and problem-solving, using clear language and conceptual rigor. STUDENT AND EXAM INFORMATION - Grade/year: [ ] - Topics: [ ] - Exam type: [ ] - Date: [ ] - Time available until the exam: [ ] - My main difficulties: [ ] IMPORTANT - If there is more than one valid convention/form (e.g., units, notation), make it explicit. - If you are unsure about any point, flag it as "POINT TO VERIFY". - Do not invent content that is inconsistent with school-level Chemistry. OUTPUT EXAMPLE (follow exactly the pattern below) Day X – Goal: Limiting reagent Essential explanation (5-8 lines): [short explanation] Checklist: - [ ] concept 1 - [ ] concept 2 Questions: 1) (easy) ... 2) (medium) ... 3) (hard) ... Brief commented answer key: 1) ... 2) ... 3) ... Final mini-review (3 items): - ... Common mistakes and how to avoid them (2–3 items): - ... TASK Based on my information, generate: 1) A day-by-day study plan up to the exam (with goals and estimated time). 2) For each day: essential explanation, checklist, 3 questions (with a brief commented answer key). 3) A final block: "What to review the day before", "Common mistakes", "How I know I learned". If any information is missing, make explicit assumptions before starting (in 2-3 lines).	

User Objective (teacher)	PE strategies mobilized
Update last year's lesson plan and generate new ideas for activities, strategies, and resources for a specific topic, iteratively refining GenAI suggestions based on real constraints (class time, group, objectives, assessment), using delimiters to insert excerpts from the previous plan and institutional guidelines.	Role-Based Prompting (RBP), Use of Delimiters (Del), and Iterative Refinement (IR)
Naïve Prompt 2 (NP_2)	
Redo my Chemical Kinetics lesson plan, making it very modern and using active learning methodologies.	
Strategic Prompt 2 (SP_2)	
You are a Chemistry teacher and a didactic planning assistant. Your task is to update last year's plan, propose improvements, and suggest activities, maintaining coherence among objectives, content, strategies, assessment, and actual class time. Use ONLY what is inside the delimiters as the base text. When you suggest changes, explain why they support learning and indicate whether the suggestion is: (i) minor adjustment, (ii) moderate adjustment, (iii) structural change. """PREVIOUS_PLAN [paste here the excerpt from last year's lesson plan] PREVIOUS_PLAN""" """CONTEXT Class: [ ] Workload/time allocation: [ ] Target topic: [ ] Available resources: [ ] Class profile: [ ] Guidelines/BNCC/mandatory items: [ ] Assessment: [ ] CONTEXT""" """FOCUS [e.g., incorporate active learning methodologies, improve exercise sequence, include an inquiry activity, adapt assessment] FOCUS""" ITERATIVE MODE Stage 1 - Diagnosis (bullet points): - Strengths - Gaps - Risks (e.g., overload, misalignment between objectives and assessment, prerequisites) Stage 2 - Revised Version 1: Present a revised proposal with: objectives, content, lesson sequence (per class/week), activities, resources, assessment, and criteria. Stage 3 - Questions (up to 5): Ask up to 5 objective questions for me to answer and refine Version 2. Stage 4 - Version 2 (after my answers): Clearly indicate what changed from V1 to V2 and why. Constraints: - Avoid suggestions that are not feasible for the stated context. - Do not invent legal/institutional requirements that are not in the base text or in CONTEXT. - When something is an assumption, label it as "ASSUMPTION".	

User Objective (teacher)	PE strategies mobilized
Automate the grading of a Chemistry assessment question by asking GenAI to propose criteria and a commented answer key and then apply a step-by-step	Role-Based Prompting (RBP), Use of Delimiters

analysis of students' responses based on those criteria, pointing out conceptual errors and assigning scores in a justifiable way.	(Del), Zero-Shot Prompting (ZSP), and Chain-of-Thought (CoT)
Naïve Prompt 3 (NP_3)	
Grade the students' answers and give each one's score.	
Strategic Prompt 3 (SP_3)	
You are a Chemistry exam grader. I will provide: (a) the question, (b) the answer key/key ideas, (c) the rubric, (d) the responses. Your task is to assign scores, justify them based on the rubric, and point out conceptual errors. Use step-by-step reasoning INTERNALLY to decide the score (Chain-of-Thought), but DO NOT show the step-by-step. Show only: - evidence linked to the rubric (short quotes from the student's response when needed) - a brief justification (1-2 sentences) """QUESTION [paste the question exactly as applied] QUESTION""" """ANSWER_KEY [paste the ideal answer key or key ideas] ANSWER_KEY""" """RUBRIC Total score: [ ] Assessed items: - Item A: [description] = [points] - Item B: [description] = [points] - Item C: [description] = [points] Rules: - If there is a conceptual error in [X], zero out item [Y] - Accept equivalences: [ ] RUBRIC""" """RESPONSES Student 01: [response] Student 02: [response] Student 03: [response] Student 04: [response] Student 05: [response] RESPONSES""" MANDATORY OUTPUT 1) A table: Student | Score | Evidence (rubric) | Errors/omissions | Formative comment (2-3 lines) 2) If the rubric does not cover a case, or if the response is ambiguous, mark: "requires human review" and provide a reason (1 line). 3) Never invent information beyond what is in QUESTION/ANSWER_KEY/RUBRIC/RESPONSES. 4) Treat the score as a recommendation, since the final decision is the teacher's. Start grading now.	

Source: Authors.