Int. j. adv. multidisc. res. stud. 2026; 6(5):75-84

Received: 13-07-2026
Accepted: 23-08-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

The Epistemological Status of AI-Generated Knowledge in Science Education: A Theoretical Examination of Knowledge, Evidence, and Trust

Dr. Özgür Özünlü
Lecturer, Health Services Vocational School, Ankara Yildirim Beyazit University, Ankara, Turkey

DOI: https://doi.org/10.62225/2583049X.2026.6.5.6828
Corresponding Author: **Dr. Özgür Özünlü**

Abstract
The increasing integration of generative artificial intelligence (GenAI) systems into educational environments has brought about not only a technological transformation of teaching and learning processes but also fundamental epistemological questions concerning the sources of knowledge, its verification and justification, and the ways in which its reliability should be evaluated. This issue is particularly salient in science education, where a central learning objective is for students to relate scientific claims to evidence, recognize uncertainty, and critically examine how explanations are constructed.
This study aims to theoretically examine the epistemological status of AI-generated outputs in science education through the concepts of scientific evidence, the Nature of Science (NOS), epistemic autonomy, epistemic agency, and human oversight. The study was not designed as an empirical document analysis or a systematic review; rather, it develops a conceptual synthesis drawing on current literature in education, science education, and artificial intelligence. The analysis argues that AI-generated outputs should not be positioned directly as scientific evidence or verified knowledge, but rather as provisional knowledge claims that require verification. To translate this epistemological positioning into pedagogical practice, the article proposes a seven-stage framework entitled the Epistemic Verification Cycle. Each stage of the cycle is theoretically justified, and an illustrative classroom implementation scenario is developed using photosynthesis as an example.
In addition, AI hallucinations and the generation of fabricated references are analyzed in a separate section as structural characteristics of epistemic risk rather than as merely marginal system failures. Overall, the value of AI in science education depends less on its ability to generate correct answers rapidly than on how it is pedagogically designed to support students ’capacities for evidence evaluation, scientific reasoning, and epistemic responsibility.

Keywords: Artificial Intelligence, Generative Artificial Intelligence, Science Education, Epistemology, Scientific Evidence, Nature of Science, Epistemic Autonomy, Hallucination, AI Literacy

1. Introduction

The rapid incorporation of generative artificial intelligence systems into educational environments is bringing about a significant transformation in processes of accessing and producing knowledge. Large language models (LLMs) can answer questions, generate explanations, summarize texts, propose alternative solutions, and engage in dialogue with users. These capabilities provide important opportunities for personalized support, rapid feedback, and the provision of alternative forms of explanation within teaching and learning processes. However, the ability of generative AI to produce fluent and persuasive texts does not necessarily imply that the generated content is scientifically accurate or evidence-based (Kasneci *et al*., 2023; Tlili *et al*., 2023) [14, 31].

The issue is even more pronounced in science education. Scientific knowledge does not consist merely of statements that appear to be correct; rather, it is constructed and continuously tested through practices such as observation, measurement, experimentation, modeling, inference, critical evaluation, and justification within scientific communities. Duschl emphasizes that science education cannot be reduced solely to the acquisition of conceptual knowledge; instead, conceptual, epistemic, and social learning goals need to be addressed together. Accordingly, whether an explanation generated by an AI system should be accepted as scientific knowledge depends not only on the accuracy of the explanation but also on the evidence upon which it is based and the manner in which the student verifies it.

One of the most fundamental risks associated with the epistemic positioning of generative AI in science education is that such systems may present themselves as ultimate epistemic authorities. As indicated in Grant Cooper's study examining ChatGPT's responses to science education questions, ChatGPT may create an impression in which "a single truth is assumed without sufficient evidence or qualification" (Cooper, 2024). This is not merely an issue of content accuracy; it directly influences students' understanding of how scientific knowledge is justified.

As Erduran emphasizes, AI is not merely a tool entering learning environments but also a force that is transforming how science itself is conducted. Science education therefore needs to reflect this transformation and critically examine what constitutes a "Nature of Science shaped by AI" and what competencies this emerging context requires from students (Erduran, 2023) [9]. Accordingly, the research questions addressed in this study extend beyond the question of whether AI produces "correct answers" to encompass students' epistemic relationships with AI-generated outputs.

Recent research indicates that generative AI can be employed to support scientific inquiry and epistemic understanding in science education. For example, Chang, Park, and Sim demonstrated that, when supported by appropriately designed prompts, GenAI has the potential to assess students' epistemic understanding of scientific inquiry and provide student-centered feedback.

Within this framework, the central question of the study is formulated as follows:

Under what conditions can AI-generated knowledge be regarded as a reliable learning resource in science education, and under what conditions can it constitute an epistemic risk?

Based on this central question, the study focuses on three sub-questions:

1. How can the relationship between AI-generated outputs and the epistemological characteristics of scientific knowledge be explained?
2. How should evidence, verification, and uncertainty be structured in AI-supported science learning?
3. What pedagogical design principles can be adopted to preserve students' epistemic autonomy?

In addressing these questions, the article seeks to address three gaps in the existing literature: (i) moving the analysis of AI hallucinations and fabricated references beyond the level of a pedagogical warning toward an epistemological examination; (ii) providing a theoretical justification for each stage of the proposed verification framework; and (iii) concretizing the classroom transfer of the framework through an illustrative implementation scenario.

2. Method of the Theoretical Review

This article is structured as a theoretical review. Its purpose is to synthesize findings from different research traditions within a single epistemological framework and to propose an explanatory model that can be applied in science education. The theoretical synthesis was conducted along three main strands of literature.

The first strand encompasses discussions concerning the opportunities and limitations of generative AI in education, issues of accuracy, cognitive offloading, and human oversight (Kasneci *et al.*, 2023 [14]; Tlili *et al.*, 2023 [31]; Miao & Holmes, 2023). The second strand consists of science education literature addressing the relationship between scientific knowledge and evidence, explanation, argumentation, and the Nature of Science (Duschl, 2008 [8]; Jiménez-Aleixandre & Erduran, 2008). The third strand comprises recent studies directly examining the relationship between GenAI and scientific inquiry, epistemic understanding, and argumentation (Tang, 2024; Tang & Cooper, 2025; Cheung & Zhang, 2026 [5]; Chang *et al.*, 2026 [3]; Deguchi *et al.*, 2026 [7]).

Accordingly, the contribution of this article is not to retest empirical findings from the literature using a new dataset, but rather to develop a conceptual framework that explains the epistemic positioning of AI-generated outputs in science education and that can be transferred to classroom practice. Consequently, the *Epistemic Verification Cycle* and the implementation scenario presented in this study are proposed as propositions to be empirically tested rather than as empirical findings.

3. Conceptual Framework

3.1 Distinguishing Between Knowledge, Evidence, and AI-Generated Outputs

Scientific knowledge and textual outputs generated by artificial intelligence cannot be placed within the same epistemic category. Large language models primarily generate probable continuations based on linguistic patterns; the fluency, detail, and persuasiveness of a response do not guarantee that it accurately represents a phenomenon in the real world. This distinction necessitates explicit instruction concerning the difference between "fluent accuracy" and "evidence-based accuracy" in the use of generative AI (Kasneci *et al.*, 2023; Tang & Cooper, 2024) [14, 28].

The epistemic strength of a scientific claim depends not merely on the existence of the claim, but also on the quality of the evidence supporting it, the justificatory relationship between the evidence and the claim, and the evaluation of alternative explanations. Therefore, an AI-generated output should initially be treated as a claim rather than as a conclusion.

Instead of accepting an AI-generated statement as inherently true, students should ask questions such as:

- What evidence supports this claim?
- What is the source?
- Does this source actually exist?
- Do other sources support the same conclusion?
- What uncertainties are associated with this conclusion?

The technical basis of this distinction lies in the generation mechanism of large language models. These models operate as automated completion systems based on statistical patterns rather than on an intrinsic understanding of facts, truth, originality, or meaning (Fouesneau *et al.*, 2024) [11]. Consequently, the grammatical and rhetorical fluency of a sentence generated by a model provides no information about whether its content corresponds to reality. The objective of the model is to predict the most probable meaningful sequence of words rather than necessarily to provide the most accurate and factual information (Sadeghpour *et al.*, 2025) [25]. This mechanism also helps explain why hallucinations and fabricated references represent structural characteristics of these systems, as discussed in subsequent sections (Temsah *et al.*, 2025) [30].

3.2 The Nature of Science and Generative AI

The Nature of Science encompasses understandings of how science produces knowledge, what characteristics scientific knowledge possesses, and the values and methods through which scientific practices are conducted. This dimension is particularly important in science education because it enables students to understand the distinction between a "correct answer" and "scientifically justified knowledge."

The relationship between GenAI and the Nature of Science is bidirectional. On the one hand, AI can assist students in generating alternative explanations, making comparisons, and reconsidering their scientific arguments. On the other hand, AI may render less visible the processes through which scientific knowledge is constructed, including observation, measurement, experimentation, modeling, and criticism within scientific communities.

Cheung and Zhang's study of GPT-4o's representations of the Nature of Science demonstrated that GenAI can reflect certain characteristics of science but may not consistently make the distinction between scientific knowledge production and AI-based knowledge generation sufficiently explicit. This finding suggests that AI outputs should also be evaluated from a Nature of Science perspective in science classrooms.

According to Erduran, this discussion should be conducted at two levels: not only should the opportunities AI provides for science—such as simulations, personalized learning, and targeted feedback—be considered, but the ways in which AI transforms the Nature of Science should also become part of the educational agenda (Erduran, 2023) [9]. Cheung and Zhang's study using GPT-4o further confirms this duality. GPT-4o was able to represent certain characteristics of science associated with the cognitive-epistemic categories of the Family Resemblance Approach; however, it demonstrated both strengths and weaknesses in distinguishing between the Nature of Science and an AI-influenced Nature of Science (Cheung & Zhang, 2026) [5]. This finding reinforces the argument that evaluating AI outputs from a Nature of Science perspective should be treated not merely as a form of monitoring, but as a component of the learning content itself.

3.3 Epistemic Autonomy and Epistemic Agency

Epistemic autonomy refers to students' capacity not to completely delegate judgments concerning what to believe and which claims to accept to another person or an algorithm. Epistemic agency, in turn, refers to students' active responsibility in asking questions, seeking evidence, constructing claims, providing justifications, and revising their views when necessary. GenAI can strengthen or weaken both dimensions.

Theoretical studies published in 2026 emphasize that learning with GenAI involves more than technological competence; it requires students to reason both with AI and against AI. Similarly, the notion of epistemic partnership proposes that AI should be positioned not as an ultimate authority but as a temporary thinking partner that can be used during processes of inquiry and reconstruction (Deguchi *et al*., 2026 [7]; Samuel, 2026).

The theoretical basis of epistemic agency derives from the following definition in the science education literature: epistemic agency concerns "the ways in which students, teachers, or social groups take up and distribute cognitive authority—that is, responsibility for creating and evaluating new knowledge" (Odden *et al*., 2021) [20].

In the age of GenAI, this definition can be extended by incorporating AI-generated outputs into the pool of "knowledge claims to be evaluated." Students should assume responsibility for verifying AI-generated content, cross-checking sources, and seeking human judgment in situations requiring interpretation, ethical consideration, and contextual understanding (Jose *et al*., 2025) [13].

Wu and colleagues conceptualize this process as shared epistemic agency between humans and GenAI, arguing that students possessing adaptive epistemic stances may derive greater benefits from GenAI (Wu *et al*., 2025) [36]. When left uncontrolled, however, the process may reinforce a pattern of "passive epistemic consumption," thereby undermining students' agency and teachers' roles as guides for critical thinking. Consequently, resisting epistemic automation becomes a pedagogical necessity (Jose *et al*., 2025) [13].

4. The Epistemological Status of AI-Generated Knowledge

The main argument of this article is that AI-generated outputs should not be positioned as "knowledge", but rather as "knowledge claims" that require verification. This is because, when generating a claim, generative AI does not conduct an experimental measurement, does not have direct access to the physical world, and cannot independently provide an epistemic guarantee of the truth of the statements it produces.

This distinction is particularly important in scientific contexts. For example, when a student encounters an AI-generated statement such as "Plants perform photosynthesis in darkness," the instructional objective should not merely be to provide the correct answer. Instead, students should be expected to treat the statement as a hypothesis, examine the conditions required for photosynthesis, identify appropriate experimental evidence, and compare information obtained from different sources. In such a context, AI ceases to function merely as a repository of information and becomes a tool that stimulates epistemic inquiry.

This perspective requires a more sophisticated evaluative approach than the binary question of whether "AI is right or wrong." An AI output may be correct; however, its correctness does not make it scientific evidence. Similarly, an output may be incorrect, but identifying its incorrectness through evidence can generate high-level scientific learning. Thus, from a pedagogical perspective, the value lies not solely in the accuracy of the output but in the epistemic relationship that students establish with that output.

4.1 Epistemic Trust: Trust or Verification?

Reducing trust in AI within educational environments to two extremes is inappropriate. At one extreme lies unconditional trust, while at the other lies complete distrust. A more functional approach is calibrated trust. Calibrated trust requires users to distinguish between tasks in which the system is strong and those in which it is weak, and to adjust their level of trust according to the quality of the available evidence.

Educational research emphasizes that critical thinking, fact-checking, and continuous human oversight are necessary for LLMs to be used beneficially. In science education, calibrated trust can be operationalized through source tracking, independent verification, experimental comparison, and uncertainty analysis. The statement "AI said so" should not be accepted as a scientific justification. Instead, students should be expected to formulate their reasoning as follows: "AI proposed this claim, and this

claim is supported by the following independent evidence."
Strong empirical evidence supports the need for calibrated trust. Steyvers *et al*. (2025) [27] experimentally demonstrated that users tend to overestimate the accuracy of LLM responses and that user confidence increases as explanations become longer, even when increased length does not improve accuracy. Similarly, Klingbeil *et al*. (2024) [15] showed that simply knowing that a recommendation was generated by AI can lead to overreliance when the recommendation conflicts with contextual information or individuals' own judgments.

Safarov *et al*. (2026) [26], in an experiment involving 248 university students, found that particularly among students without domain expertise, the intuition that "AI knows better" may operate readily; polished presentation can easily be interpreted as a sign of reliability, increasing the likelihood of responses that align with incorrect recommendations. Their findings indicate that trust calibration cannot be achieved through a single cue; rather, student performance, AI reliability, and explanation design need to be considered together. Accordingly, independent reasoning before consulting AI and brief verification procedures before accepting AI-generated information should be incorporated into instruction.

Undertrust also constitutes an independent problem. Walker *et al*. (2024) [34], in an experiment involving 171 students, found that students did not excessively trust ChatGPT but demonstrated undertrust in AI when responses were explicitly labeled as AI-generated, reflecting a pro-human bias. They emphasized that, just as overtrust can be problematic, undertrust can deprive students of valuable learning opportunities. Pal *et al*. (2026) [23] further demonstrated that trust and reliance behaviors may diverge in educational contexts: students may have high trust in human experts but consult them less frequently because of concerns about being judged, while they may report lower trust in AI systems but rely on them extensively because of accessibility, anonymity, and the absence of perceived judgment. This distinction suggests that trust in science education should be monitored not only at the attitudinal level but also behaviorally, particularly through source-selection and verification routines.

4.2 Materiality and Scientific Evidence

An important characteristic of scientific knowledge is its relationship with the material world. Observation, measurement, experimental arrangements, specimens, instruments, and data all contribute to the construction of scientific claims. Tang and Cooper emphasize the importance of materiality in the age of GenAI and argue that textual production cannot completely replace scientific relationships established with the physical world. Accordingly, in science education, AI-generated outputs do not possess the same epistemic status as experimental data.

This distinction is particularly important in laboratory and inquiry-based activities. For example, an AI system may explain the expected outcome of a chemical reaction; however, the measurement obtained by students in an actual experiment, the experimental conditions, and the sources of error possess a different epistemic status. An AI-generated explanation should therefore assist in interpreting experimental findings rather than replace them.

Tang and Cooper (2024) [28] emphasize that scientific authority derives from our relationship with the physical world and argue that materiality will become an increasingly important criterion for evaluation in science education.

4.3 Hallucinations and Fabricated References: The Structural Nature of Epistemic Risk

Discussions of the epistemic risks associated with AI often remain at the level of the statement that "AI sometimes makes mistakes." However, when the generation mechanism described in Section 3.1 is taken into account, hallucination—that is, the production of fabricated, inconsistent, or non-referential content that does not correspond to the real world—should be understood as a structural characteristic rather than simply a system malfunction.

Because large language models probabilistically generate linguistic continuations that appear most plausible, the fluency and confident tone of an output provide no epistemic information regarding its accuracy (Kasneci *et al*., 2023; Tlili *et al*., 2023) [14, 31].

This structural characteristic manifests itself in science education in three primary ways:

1. **Content-level inaccuracies:** The model may generate scientifically incorrect but pedagogically persuasive explanations. Because such outputs may present misconceptions in the same rhetorical style as scientifically accurate explanations, students may be unable to distinguish correct and incorrect information through superficial reading.
2. **Fabrication at the source level:** The model may generate nonexistent studies, authors, or DOIs. Consequently, source verification, discussed in Stage 2, must be transformed from an optional habit into a mandatory epistemic procedure. The existence of an AI-provided reference must be independently verified, and its actual support for the claim must be examined separately.
3. **Concealment of uncertainty:** Indicators of uncertainty in scientific texts, such as measurement error, sample limitations, and alternative explanations, may disappear within the fluent and definitive language generated by the model. This constitutes an epistemic distortion that can weaken students' understanding of the provisional nature of scientific knowledge.

This analysis produces two implications. First, verification of AI-generated outputs in science education is not merely a safety precaution; it is a natural component of teaching the Nature of Science. When students encounter a hallucinated statement, they experience concretely why an apparently plausible statement does not necessarily constitute scientific knowledge. Second, the risk of hallucination requires verification throughout the entire cycle. Verification is not an action completed at a single stage but a process extending from the source of the claim to the student's personal judgment.

Empirical evidence concerning the magnitude of this structural risk is increasing. According to findings reported by Chen and Chen. (2024), 98% of references generated by ChatGPT-3.5 were fabricated, whereas this proportion decreased to 20% for ChatGPT-4. Nevertheless, the fundamental issue that the model does not genuinely "know" information means that the problem is not completely eliminated even in models connected to the internet. Linardon *et al*. (2025) [17] reported that approximately two-thirds of references generated by GPT-

4o were fabricated or erroneous, with the rate varying according to topic familiarity and prompt specificity. Mugaanyi *et al*. (2024) [19] demonstrated methodologically that references generated by ChatGPT need to be independently verified through Google Scholar, PubMed, and DOI databases.

Temsah *et al*. (2025) [30] explain that such "well-formatted but fictitious" references may result from the model's inability to access subscription-based databases and from its probabilistic text-generation mechanism. Sadeghpour *et al*. (2025) [25] likewise emphasize that models predict the most probable meaningful sequence of words rather than necessarily generating the most accurate factual information. Consequently, the use of AI-generated outputs in academic research without verification and transparency regarding AI contributions is clearly problematic.

5. The Epistemic Verification Cycle: A Proposed Framework for Science Education

When the central discussions in the literature are considered together, the safe and pedagogically productive use of GenAI in science education can be structured within a verification cycle. The framework proposed below constitutes the theoretical contribution of this article.

The design logic of the cycle derives from two principles: (i) as justified in Section 4, an AI-generated output should always be treated as a knowledge claim; and (ii) as analyzed in Section 4.3, verification should be maintained throughout the process rather than being confined to a single stage.

5.1 Stage 1 — Identifying the Claim

Students should read an AI-generated output not as a unified text but as a collection of distinct scientific claims. The purpose of this stage is to make the propositions embedded within fluent discourse visible: Which statements contain causal claims? Which are descriptions? Which are generalizations? This stage directly operationalizes the positioning of AI output as a "knowledge claim" discussed in Section 4 and corresponds to the development of scientific claim and hypothesis-formulation skills.

5.2 Stage 2 — Tracing the Source

Students trace the source, data, and references underlying the claim. As shown in Section 4.3, AI systems can generate nonexistent sources; therefore, source verification at this stage involves two levels: (i) independently verifying whether the cited source actually exists and (ii) examining whether the source genuinely supports the claim. This stage activates source literacy and understanding of the hierarchy of evidence.

5.3 Stage 3 — Comparing Evidence

Students compare the claim with at least one independent source, such as a peer-reviewed study, textbook, or institutional report, and/or with their own experimental data. In accordance with the discussion of materiality in Section 4.2, experimental data possess a higher epistemic status than textual output in this comparison. Measurements obtained in the laboratory provide the basis against which the AI-generated textual explanation can be tested.

5.4 Stage 4 — Justification

Students explicitly establish the relationship between the claim, evidence, and warrant. A Toulmin-based scientific argumentation approach provides the structural framework for this stage. The critical point is that the justification should not refer to AI as the authority; AI's role is limited to having generated or introduced the initial claim.

5.5 Stage 5 — Evaluating Uncertainty

Students identify the limitations of the conclusion they have reached and consider alternative explanations: Under what conditions is the conclusion valid, and under what conditions might it change? This stage addresses the risk of "concealed uncertainty" identified in Section 4.3 and incorporates an understanding of the provisional nature of scientific knowledge into the learning process.

5.6 Stage 6 — Personal Judgment

Students formulate their own conclusions independently of AI and justify the evidence upon which those conclusions are based. This stage constitutes the operational manifestation of epistemic autonomy defined in Section 3.3. Evaluation focuses not on whether the AI-generated content was correct but on how students justify the claim.

5.7 Stage 7 — Reflection

Students evaluate the contribution and limitations of AI to their own thinking processes: At what point was AI helpful? At what point was it misleading? To what extent did I deviate from my initial idea? This stage develops epistemic awareness and transforms calibrated trust—understood as the ability to distinguish between tasks for which the system is strong and weak—into a learning outcome.

The distinctive characteristic of the cycle is that it transforms AI use from an activity aimed at "finding the answer" into an activity aimed at "determining the scientific status of a claim."

The design logic of the cycle is also consistent with recent pedagogical frameworks developed for LLM integration in higher education. Vendrell and Johnston (2026) [33] argue that unstructured LLM use may weaken critical thinking through cognitive offloading, metacognitive disengagement, and loss of epistemic agency. In contrast, they recommend positioning LLMs as temporary thinking partners, maintaining assessment throughout learning, and designing AI-mediated and AI-free stages sequentially.

Table 1: Student Actions, Core Epistemic Questions, and Science Education Outcomes

S. No	Stage	Student Action	Core Epistemic Question	Science Education Outcome
1	**Identifying the Claim**	Identifies the central claim in the AI-generated output	What exactly is AI claiming?	Formulation of scientific claims/hypotheses
2	**Tracing the Source**	Independently verifies the existence and content of the source	Where does this claim come from? Does the source actually exist?	Source literacy
3	**Comparing Evidence**	Compares the claim with independent sources and/or experimental data	Does the evidence support the claim?	Evidence evaluation
4	**Providing Justification**	Establishes the relationship among the claim, evidence, and warrant	Why did I reach this conclusion?	Scientific argumentation
5	**Evaluating Uncertainty**	Identifies limitations and alternative explanations	Under what conditions might the conclusion change?	Understanding of scientific uncertainty
6	**Forming an Independent Judgment**	Expresses their own conclusion independently of AI	What evidence do I base my belief on?	Epistemic autonomy
7	**Reflection**	Evaluates AI's contribution and limitations	How did AI influence my thinking?	Epistemic awareness

6. Implementing the Cycle in the Classroom: An Illustrative Scenario

To concretize the classroom transfer of the framework, the following example demonstrates how the seven stages may be implemented using the scenario introduced in Section 4. This scenario is not an empirical finding but a design example illustrating how the framework may be interpreted pedagogically.

During preparation for a biology unit on photosynthesis, a student asks an AI tool a question and receives the response: "Plants perform photosynthesis in darkness."

- **Identifying the claim:** The student identifies the central claim in the output—"photosynthesis can occur in darkness"—as a causal generalization and formulates it as a hypothesis.
- **Tracing the source:** The student independently searches for the reference provided by the AI and verifies whether the source exists and whether it actually supports the claim. Even if the source proves to be fabricated, the process does not stop; this finding becomes material for reflection in Stage 7.
- **Comparing evidence:** The student tests the claim against the textbook's explanation of the conditions required for photosynthesis and, where possible, through a simple experimental arrangement comparing plant samples exposed to light and darkness. The experimental observation possesses a higher epistemic status than the AI-generated textual claim.
- **Justification:** The student relates evidence demonstrating the necessity of light for photosynthesis to the claim and constructs an argument using the structure of "claim–evidence–warrant." The statement "AI said so" cannot function as a valid justification.
- **Evaluating uncertainty:** The student investigates the boundaries of the conclusion: Which stages of photosynthesis are light-dependent? Do some reactions continue in darkness? Thus, an initially simplistic and incorrect statement is transformed into a scientifically more nuanced question.
- **Personal judgment:** Independently of the AI output, the student formulates the conclusion: *"Light is necessary for the light-dependent stages of photosynthesis,"* together with the supporting evidence.
- **Reflection:** The student reflects on how the AI response could be both incorrect and persuasive, what questions the response prompted, and how the student should approach AI-generated outputs in the future.

This scenario operationalizes the central argument of Section 4: when appropriately structured, an incorrect AI output can serve as a starting point for high-level scientific learning.

7. Pedagogical Implications for Science Education

7.1 Positioning AI Outputs as "Temporary Thinking Partners" Rather Than "Final Answers"

Teachers should design AI use not as an automation mechanism through which students delegate their tasks, but as a dialogic environment in which students test their own thinking. Students should first formulate their own predictions or explanations and subsequently compare them with AI-generated suggestions. Rather than eliminating cognitive effort, this design transforms AI into a resource for comparison and critique.

This positioning is consistent with Vendrell and Johnston's (2026) [33] conceptualization of the LLM as a "temporary thinking partner": the LLM should not function as the authority making students' final decisions but as a partner temporarily involved at specific stages of the thinking process while preserving productive cognitive friction.

7.2 Teaching Source Verification and the Hierarchy of Evidence

Simply telling students that "AI can make mistakes" is insufficient. Students need to know which sources can be used to verify a claim. The distinct epistemic functions of peer-reviewed studies, scientific reports produced by universities or public institutions, textbooks, primary experimental data, and observational findings should be explicitly taught. As demonstrated in Section 4.3, sources provided by AI must also be independently verified.

7.3 Placing Scientific Argumentation at the Center of AI Use

Scientific argumentation provides a powerful pedagogical tool for evaluating the epistemic status of AI-generated outputs. When students establish relationships among claims, evidence, warrants, and counter-evidence, they can more clearly understand why an AI-generated sentence cannot, by itself, constitute scientific knowledge. Toulmin-based argumentation approaches can be used for this purpose.

Recent findings concerning the integration of AI and scientific argumentation are both promising and warrant caution. Tang and Putra (2025) [29], drawing on Bakhtin's theory of heteroglossia, demonstrated that their customized chatbot, "Dialogic Science Teacher," stimulated perspective-taking, reasoning, argumentation, and creative thinking in dialogues involving 21 students from two high schools on socioscientific issues. Their findings suggest that GenAI can be designed as a "dialogic partner rather than an authoritative information provider."

Chang *et al.* (2026) [3] reported high agreement between AI and human raters in the evaluation of 560 responses collected from 80 fourth-grade students using ChatGPT-4o (κ = 0.825), while the generated feedback strongly supported student agency (mean = 2.75/3). These findings suggest that, when appropriately designed, GenAI can function as a dialogic partner supporting students' epistemic understanding.

Watts *et al.* (2025) [35] developed principles for designing chatbots based on prompt engineering to support scientific reasoning and argumentation among middle school students. Turós *et al.* (2025) [32], using the Toulmin model to analyze students' arguments with AI, demonstrated that establishing claim–evidence–warrant connections requires instructional support.

At the same time, the quasi-experimental study by Resi *et al.* (2026) [24] showed that AI-supported debates among 72 tenth-grade students produced a significant difference only in initial elements such as "claim and data" ($p = .05$), while no significant effects were observed in other dimensions of argument structure and quality. This finding emphasizes that AI use alone is insufficient and that targeted instruction is necessary.

Lee *et al.*'s findings regarding HASbot further demonstrated that AI-based automated scoring and feedback significantly improved scientific argumentation involving uncertainty among 343 middle and high school students ($ES = 1.52$, $p < .001$). This finding suggests that the uncertainty-evaluation component of Stage 5 may be effectively supported through technological tools.

7.4 Combining AI-Supported and AI-Free Stages

Conducting every task with AI may render students' independent reasoning invisible. Therefore, instructional design can employ a staged structure involving an AI-free beginning, comparison with AI, and AI-free final justification. Such a design allows students to identify differences among their prior knowledge, AI suggestions, and scientific evidence.

The sequential use of AI-supported and AI-free stages can also be justified through the literature on cognitive offloading. Fan *et al.* (2024) [10] argue that excessive AI use may lead students toward "metacognitive laziness" through cognitive offloading. They further suggest that cognitive difficulty or disfluency can stimulate analytical thinking and that AI may inhibit deeper metacognitive processes by eliminating such productive difficulty.

Özer *et al.* (2025) [22] emphasize that delegating basic tasks to AI may weaken memory retention, reduce active engagement with content, and gradually undermine critical thinking, questioning, and source-evaluation abilities. Similarly, Zhai *et al.* (2024) [38] found that overreliance on AI dialogue systems may disrupt critical thinking, whereas activities combining AI with traditional learning materials can encourage students to critically examine AI responses, identify biases, and construct balanced arguments.

Ongesa (2026) [21] summarizes this debate by emphasizing that the distinction between cognitive offloading and cognitive augmentation depends not on the technology itself but on pedagogical design and student engagement. Therefore, the sequential structure proposed here is not merely an instructional preference but a design requirement for preserving the cognitive integrity of learning.

7.5 Redefining the Role of the Teacher

The increasing prevalence of GenAI does not diminish the importance of teachers; rather, it strengthens their role as epistemic guides. Teachers should go beyond determining which AI tools are to be used and guide students in evaluating claims and evidence, recognizing uncertainty, and constructing scientific justifications. UNESCO's human-centered approach similarly emphasizes that GenAI use in education should be regulated in ways that are ethical, safe, equitable, and pedagogically meaningful.

8. Proposed Pedagogical Design Principles

- **Principle of epistemic responsibility:** Student performance should be evaluated primarily according to how students verify claims rather than according to the content generated by AI.
- **Principle of evidence priority:** AI-generated output should not be elevated to the status of reliable knowledge without independent scientific evidence.
- **Principle of transparency:** Students should explicitly indicate at which stage and for what purpose AI was used.
- **Principle of cognitive effort:** AI use should not eliminate students' responsibility to think, explain, and justify.
- **Principle of comparison:** AI responses should be compared with at least one independent scientific source, dataset, observation, or experimental finding.
- **Principle of uncertainty:** Students should address not only whether a claim is "true or false," but also "to what extent is it reliable and under what conditions is it valid?"
- **Principle of human oversight:** Scientific, ethical, and high-stakes decisions should not be delegated to AI.
- **Principle of reflection:** Students should evaluate how AI use influenced their own ways of thinking.

These principles constitute the evaluative dimension of the Epistemic Verification Cycle. Student performance can therefore be assessed not according to the accuracy of the AI output but according to the epistemic rigor with which students implement the stages of the cycle.

The principle of epistemic responsibility can also be justified through Ongesa's (2026) [21] concept of epistemic vigilance. Because generative systems can produce fabricated references, subtle inaccuracies, and cultural biases, detecting such problems requires analytical attention.

The principle of human oversight is directly supported by UNESCO's framework. UNESCO's humanistic framework for AI in education prioritizes human control over AI and requires applications to be designed ethically, transparently, non-discriminatorily, and in an accountable manner (Chan, 2023) [1].

The UNESCO guidance on generative AI in education likewise advocates a human-centered approach based on robust regulatory frameworks and integrated teacher training (Chan & Colloton, 2024) [2]. In addition, Xia *et al.* (2024) [37] emphasize that human judgment and critical thinking cannot be replaced in the GenAI era, that AI literacy has become essential for both teachers and students, and that conventional assessment policies need to be redesigned.

9. Discussion and Conclusion

The epistemological status of AI-generated content in science education should be evaluated independently of how fluent or persuasive that content may appear. A GenAI output is neither direct scientific evidence nor verified knowledge; rather, it is a provisional knowledge claim that requires verification, sourcing, and connection to scientific evidence. Accordingly, the central pedagogical goal in science education should not be to encourage students to accept AI-generated answers, but to guide them in investigating the epistemic grounds on which such answers may or may not be accepted.

This perspective redefines the role of AI in education from instrumental automation toward epistemic partnership. However, such partnership should not be interpreted as an equal distribution of epistemic authority. Ultimate epistemic responsibility should remain with students, teachers, and the verification practices of the scientific community.

The value of AI lies not in eliminating students' responsibility for thinking but in supporting them to ask more questions, seek evidence, compare alternative explanations, and construct their own scientific justifications.

Consequently, the success of GenAI use in science education should be determined not by how extensively the technology is used, but by the extent to which students develop scientific reasoning, evidence evaluation, and epistemic autonomy while interacting with AI. The proposed Epistemic Verification Cycle provides an initial framework for transferring this objective into classroom practice. Empirically testing this framework across different age groups and science learning environments represents an important direction for future research.

Taken together, the literature indicates that the epistemological position of GenAI in science education should be distinguished between its role as a "source of information" and as an "epistemic tool." GenAI can provide students with alternative explanations, questions, examples, and opportunities for comparison; however, the scientific status of these outputs cannot be determined until they have undergone independent processes of evidence evaluation and justification.

Therefore, an important educational objective is not merely to improve the accuracy of AI systems but also to develop students' epistemic judgment regarding AI-generated outputs. This approach moves beyond the dichotomy between completely rejecting AI and adopting it unconditionally.

Studies published in 2026 indicate that appropriately designed GenAI use may be beneficial for scientific inquiry, epistemic feedback, and argumentation (Chang *et al*., 2026; Deguchi *et al*., 2026) [3, 7]. Nevertheless, this positive potential does not imply that AI automatically becomes an epistemic partner. The design of the learning environment, the structure of tasks, teacher guidance, and verification routines remain decisive variables.

In this context, epistemic trust is a broader concept than technical reliability. The fact that a system performs a particular task with high accuracy does not justify students' trusting all of its scientific claims. Trust must be calibrated according to the task and the evidence. Furthermore, as detailed in Section 4.3, because hallucinations and fabricated references are structural characteristics of these systems, calibrated trust should never be based solely on the observation that "the model is generally accurate." Trust must be established through evidence for each individual claim.

The relationship between scientific knowledge and the material world should likewise be distinguished from AI-generated textual production. Laboratory activities, observation, measurement, and data analysis should therefore not be replaced by AI-generated explanations; rather, they should serve as the epistemic ground against which AI outputs are tested.

When Cooper's (2023) [6] warning concerning AI as an "ultimate epistemic authority" is combined with Erduran's (2023) [9] call to recognize that AI is transforming the nature of knowledge production, the role of science education becomes clearer. Its task is not simply to teach students to accept or reject AI outputs, but to teach the processes through which the epistemic status of those outputs can be determined.

The Epistemic Verification Cycle, proposed as the original theoretical contribution of this study, requires students to treat AI outputs as hypotheses or claims and to process them through stages involving evidence, sources, argumentation, uncertainty, and personal justification. In this way, AI use is transformed from an activity of "finding the answer" into an activity of "determining the scientific status of a claim." This transformation is directly consistent with the epistemic objectives of science education.

10. Limitations

The primary limitation of this study is that the proposed framework has not been tested through a direct classroom experiment or empirical student data. Accordingly, the Epistemic Verification Cycle currently remains a theoretical model. The implementation scenario presented in Section 6 is a design example and does not provide evidence concerning the effectiveness of the cycle, the appropriate duration or intensity of its stages, or its feasibility for different student groups.

Second, because GenAI systems are developing rapidly, conclusions concerning the performance of a particular model should not be regarded as universally or permanently applicable. Third, epistemic trust and AI literacy may vary according to students' age, domain knowledge, and task type. Therefore, the proposed principles need to be empirically tested across different educational levels.

11. Suggestions for Future Research

Future research may focus on the following issues:

- Experimentally testing the Epistemic Verification Cycle at middle school, high school, and university levels.
- Comparing the scientific argumentation and evidence-evaluation performance of students who use AI with those who do not.
- Measuring students abilities to identify errors, uncertainties, and fabricated references in AI-generated outputs.
- Examining science teachers' knowledge and pedagogical competencies concerning the epistemological status of AI.
- Comparing different AI models in terms of their representations of the Nature of Science and scientific evidence.

- Longitudinally investigating students' epistemic agency in AI-supported laboratory and inquiry-based learning environments.
- Examining the effects of AI-supported and AI-free assessment stages on students' cognitive load, metacognition, and scientific reasoning.
- In light of findings indicating that AI-supported argumentation is effective only for certain components (Resi *et al*., 2026) [24], investigating which stages of the Epistemic Verification Cycle produce the greatest gains for particular student profiles.
- Investigating the correspondence between behavioral indicators of trust calibration, such as source selection and verification frequency, and attitudinal indicators, and identifying situations in which trust and reliance behavior diverge (Pal *et al*., 2026) [23].

12. References

1. Chan CKY. A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education. 2023; 20(1). Doi: https://doi.org/10.1186/s41239-023-00408-3
2. Chan CKY, Colloton T. Generative AI in Higher Education, 2024. Doi: https://doi.org/10.4324/9781003459026
3. Chang J, Park J, Sim JY. Harnessing Generative AI to Facilitate Epistemic Understanding of Students in Scientific Inquiry. Journal of Science Education and Technology, 2026. Doi: https://doi.org/10.1007/s10956-026-10298-5
4. Chen A, Chen DO. Accuracy of chatbots in citing journal articles. JAMA Netw Open. 2023; 6. Doi: 10.1001/jamanetworkopen.2023.27647
5. Cheung KKC, Zhang W. Representations of the Nature of Science in Generative AI (GPT-4o). Science & Education, 2026. Doi: https://doi.org/10.1007/s11191-025-00718-0
6. Cooper G. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. Journal of Science Education and Technology. 2023; 32(3):444-452. Doi: https://doi.org/10.1007/s10956-023-10039-y
7. Deguchi A, Tsuji H, Kitazawa T. Designing Epistemic Partnerships with Generative AI: A Four-Stage Model of Student Engagement in Scientific Argumentation. Res Sci Educ, 2026. Doi: https://doi.org/10.1007/s11165-026-10370-5
8. Duschl R. Science Education in Three-Part Harmony: Balancing Conceptual, Epistemic, and Social Learning Goals. Review of Research in Education. 2008; 32(1):268-291.
9. Erduran S. AI is transforming how science is done. Science education must reflect this change. Science. 2023; 382(6677). Doi: https://doi.org/10.1126/science.adm9788
10. Fan Y, Tang L, Le H, Shen K, Tan S, Zhao T, *et al.* Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology. 2024; 56(2):489-530. Doi: https://doi.org/10.1111/bjet.13544
11. Fouesneau M, Momcheva IG, Chadayammuri U, Demianenko M, Dumont A, Hviding RE, *et al*. What is the Role of Large Language Models in the Evolution of Astronomy Research? arXiv, 2024. Doi: https://doi.org/10.48550/arxiv.2409.20252
12. Jiménez-Aleixandre MP, Erduran S. Argumentation in Science Education: An Overview. In: Erduran, S., Jiménez-Aleixandre, M.P. (eds) Argumentation in Science Education. Science & Technology Education Library, vol 35. Springer, Dordrecht, 2007. Doi: https://doi.org/10.1007/978-1-4020-6670-2_1
13. Jose B, Jose B, Cleetus A, Joseph B, Joseph L, Jose B, *et al*. Epistemic authority and generative AI in learning spaces: Rethinking knowledge in the algorithmic age. Frontiers in Education. 2025; 10. Doi: https://doi.org/10.3389/feduc.2025.1647687
14. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, *et al*. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences. 2023; 103, Article ID: 102274. Doi: https://doi.org/10.1016/j.lindif.2023.102274
15. Klingbeil A, Grützner C, Schreck P. Trust and reliance on AI-An experimental study on the extent and costs of overreliance on AI. Computers in Human Behavior. 2024; 160:108352. Doi: https://doi.org/10.1016/j.chb.2024.108352
16. Lee H, Pallant A, Pryputniewicz S, Lord T, Mulholland M, Liu OL. Automated text scoring and real-time adjustable feedback: Supporting revision of scientific arguments involving uncertainty. Science Education. 2019; 103(3):590-622. Doi: https://doi.org/10.1002/sce.21504
17. Linardon J, Jarman HK, McClure Z, Anderson C, Liu C, Messer M. Influence of topic familiarity and prompt specificity on citation fabrication in mental health research using large language models: Experimental study. JMIR Mental Health. 2025; 12. Doi: https://doi.org/10.2196/80371
18. Mouta A, Sánchez EMT, Llorente AMP. Design of a future scenarios toolkit for an ethical implementation of artificial intelligence in education. Education and Information Technologies. 2023; 29(9):10473-10498. Doi: https://doi.org/10.1007/s10639-023-12229-y
19. Mugaanyi J, Cai L, Cheng S, Lu C, Huang J. Evaluation of large language model performance and reliability for citations and references in scholarly writing: Cross-disciplinary study. Journal of Medical Internet Research. 2024; 26. Doi: https://doi.org/10.2196/52935
20. Odden TOB, Silvia DW, Malthe-Sørenssen A. Using computational essays to redistribute epistemic agency in undergraduate science. arXiv, 2021. Doi: https://doi.org/10.48550/arxiv.2108.13080
21. Ongesa T. Double Intelligence: Repositioning human intelligence in the age of generative AI for teaching, learning, and assessment across diverse higher education contexts-A narrative review. F1000Research. 2026; 15:497. Doi: https://doi.org/10.12688/f1000research.179104.1
22. Özer M, Tanberkan H, Perc M. Artificial Intelligence Threatens Critical Thinking in Education Systems. Journal of Higher Education and Science. 2025; 15(2):157-164. Doi: https://doi.org/10.5961/highereducsci.1747885
23. Pal O, Duddu V, Goyal A, Goel D, Saha K. Do We Know What They Know We Know? Calibrating

Student Trust in AI and Human Responses Through Mutual Theory of Mind. arXiv, 2026. Doi: https://doi.org/10.48550/arxiv.2601.16960
24. Resi DN, Sutarno S, Indrowati M. Exploring the impact of debates supported by artificial intelligence: Innovative strategies to empowering scientific argumentation skills in biology classes. EDUPROXIMA Jurnal Ilmiah Pendidikan IPA. 2026; 8(1):110-121. Doi: https://doi.org/10.29100/.v8i1.8172
25. Sadeghpour M, Arakala A, Rao A. Generative AI in collaborative academic report writing: Advantages, disadvantages, and ethical considerations. arXiv, 2025. Doi: https://doi.org/10.48550/arxiv.2504.08832
26. Safarov C, Gadzinski G, Schlögl S. In AI we trust? Exploring the role of explainable GenAI and expertise in education. Human Factors. 2026; 68(4):487-503. Doi: https://doi.org/10.1177/00187208251401773
27. Steyvers M, Tejeda H, Kumar A, Belém C, Karny S, Hu X, *et al*. What large language models know and what people think they know. Nature Machine Intelligence. 2025; 7(2):221-231. Doi: https://doi.org/10.1038/s42256-024-00976-7
28. Tang K, Cooper G. The role of materiality in an era of generative artificial intelligence. Science & Education. 2024; 34(2):731-746. Doi: https://doi.org/10.1007/s11191-024-00508-0
29. Tang K, Putra GBS. Generative AI as a dialogic partner: Enhancing multiple perspectives, reasoning, and argumentation in science education with customized chatbots. Journal of Science Education and Technology, 2025. Doi: https://doi.org/10.1007/s10956-025-10240-1
30. Temsah M, Al-Eyadhy A, Jamal A, Alhasan K, Malki KH. Authors reply: Citation accuracy challenges posed by large language models. JMIR Medical Education. 2025; 11. Doi: https://doi.org/10.2196/73698
31. Tlili A, Shehata B, Adarkwah MA, *et al*. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learn. Environ. 2023; 10:15. Doi: https://doi.org/10.1186/s40561-023-00237-x
32. Turós M, Kenyeres AZ, Balla G, Gazdag E, Szabó ER, Szűts Z. A Toulmin model analysis of student argumentation on artificial intelligence. Education Sciences. 2025; 15(9):1226. Doi: https://doi.org/10.3390/educsci15091226
33. Vendrell M, Johnston S-K. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. Computers and Education: Artificial Intelligence. 2026; 10:100572. Doi: https://doi.org/10.1016/j.caeai.2026.100572
34. Walker F, Favetta M, Hasker L, Walker R. They prefer humans! Experimental measurement of student trust in ChatGPT, 2024, 1-7. Doi: https://doi.org/10.1145/3613905.3650955
35. Watts FM, Liu L, Ober T, Yi S, Valle EJ-D, Zhai X, *et al*. A framework for designing an AI chatbot to support scientific argumentation. Education Sciences. 2025; 15(11):1507. Doi: https://doi.org/10.3390/educsci15111507
36. Wu J, Lee Y, Chai CS, Tsai C. Strengthening human epistemic agency in the symbiotic learning partnership with generative artificial intelligence. Educational Researcher. 2025; 54(6):358-368. Doi: https://doi.org/10.3102/0013189x251333628
37. Xia Q, Weng X, Ouyang F, Lin T, Chiu TKF. A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education. 2024; 21(1). Doi: https://doi.org/10.1186/s41239-024-00468-z
38. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students cognitive abilities: A systematic review. Smart Learning Environments. 2024; 11(1). Doi: https://doi.org/10.1186/s40561-024-00316-7