

The nature of science in the age of generative artificial intelligence: Epistemic authority, materiality, and a synergetic response

Zarifa Odilovna Muminova^{1*} , Dilmurod Mamadaliyevich Mamatkulov² ,
Rano Ibrogimovna Turokulova¹ , Nigina Tolibovna Rakhmonova¹ ,
Oybek Salimboy O'g'li Shavqiyev¹ , Sojida Umbarovna Zoirova¹ 

¹ Samarkand State University named after Sharof Rashidov, Samarkand, UZBEKISTAN

² Samarkand State Pedagogy Institute, Samarkand, UZBEKISTAN

Received 11 July 2026 • Accepted 20 September 2026

Abstract

Generative artificial intelligence has entered science classrooms faster than the epistemological questions it raises have been addressed. This study is an attempt to reveal how the knowledge-producing character of text-based conversational systems diverges from institutional science, and what that divergence implies for teaching the nature of science in secondary and undergraduate science. The two are compared through the family resemblance approach across aims and values, methods, practices, knowledge, and social-institutional systems, and the procedure linking literature to judgement is reported. The comparison is then reframed through synergetic: science is an open system in continuous evidential exchange with material reality, whereas the model reorganizes a corpus closed to new evidence at inference. A coupled-system model follows: the classroom verification norm is the order parameter, the teacher-set verification requirement the control parameter, and materiality the boundary condition. Five propositions are derived, the last a path-dependent threshold prediction that can falsify the model.

Keywords: nature of science, generative artificial intelligence, epistemic authority, materiality, synergetic, family resemblance approach

INTRODUCTION

Generative artificial intelligence (GenAI) tools have entered science classrooms at a pace rarely seen with previous educational technologies. Chat tools based on large language models have spread to nearly every area of the science curriculum within two years of their public release. Students have come to use these tools to explain concepts, draft laboratory reports, and answer content-related questions (Yan et al., 2024). This rapid proliferation has left science education research in a reactive position. Hence, it is not surprising to see that the early literature focused on practical questions such as usage, benefits, and academic integrity, while epistemological questions remained in the background. Avraamidou (2024, p. 2573), in a commentary assessing the wave of AI in science education, pointed to its lack of a theoretical foundation, describing it as “it lacks vision,

it is atheoretical, it is decontextualized”. A recent systematic review reaches a compatible conclusion from empirical material, reporting that studies of GenAI in science education have concentrated on students’ understanding either of science or of the tool, and rarely on the relation between the two (Cheung et al., 2025b).

A clarification regarding terminology and scope is necessary. In this article, “GenAI” refers to the family of technologies that generate text and similar content. A generative model, in turn, refers to the individual system with which the student actually interacts, that is, a chat tool based on a large language model. Where the context is clear, the generative model is referred to simply as the “model”. For the theoretical framework proposed in this study, only the terms “coupled-system model” and “proposed model” are used; where the concept of a model appears in the context of scientific modelling practice, it is designated as a “scientific model”. The

Contribution to the literature

- This study responds to the scarcity of work on how AI-generated knowledge relates to scientific knowledge in science education. It offers a systematic family resemblance approach (FRA) comparison of institutional science and text-based conversational systems across aims and values, methods, practices, knowledge, and social-institutional systems.
- It identifies openness to material feedback as the common dimension behind the differences in all five categories. Using synergetic, it contrasts science as an open system with a generative model that is operationally closed to new evidence at inference. It proposes a coupled-system model in which the classroom verification norm is the order parameter, the teacher-set verification requirement is the control parameter, and materiality is the boundary condition. This moves the debate beyond adoption versus prohibition toward concrete design decisions.
- It derives five empirically testable propositions, the last being a falsifiable prediction of threshold behavior and hysteresis in the classroom verification norm. This distinguishes the framework from existing norm-, scaffolding-, and epistemic-cognition-based accounts and gives the field a clear research agenda.

scope of the argument is deliberately restricted in two respects. The educational settings addressed are secondary and undergraduate science, where independent investigation and argumentation are curricular expectations. The systems addressed are text-based conversational systems as they are encountered by learners. Configurations equipped with retrieval and tool use are treated as a boundary case, whereas image and video generators, and autonomous experimentation platforms in which an AI system directs and executes laboratory work, fall outside the scope for reasons set out.

Two epistemological questions form the starting point of this study. The first question concerns epistemic authority, and the concept requires more careful handling than it usually receives. Three situations are distinguished here. Warranted epistemic dependence refers to justified reliance on a source whose claims can be traced, whose errors can be estimated, and whose authority can be contested; a student's reliance on a teacher or on a peer-reviewed report is of this kind. Perceived authority refers to the standing a source acquires through surface features such as fluency, immediacy, and apparent certainty. Uncritical deference refers to the suspension of any demand for justification. Only the third situation is at odds with the aims of science education, and the risk identified in the literature concerns the slide from the second to the third. Cooper (2023, p. 444) has highlighted the risk of a conversational model positioning itself as the "ultimate epistemic authority" in science education. Ferrario et al. (2024) argue on independent grounds that genuine epistemic expertise cannot be attributed to such systems, since expertise requires a relation to understanding that these systems do not satisfy. Empirical work points in the same direction from the side of the learner; a systematic review has reported that students' overreliance on AI dialogue systems can weaken their decision-making, critical thinking, and analytical reasoning (Zhai et al., 2024). The vigilance advocated in this article is therefore

calibrated trust rather than generalized distrust, and it is compatible with the trust that teachers and scientific institutions legitimately command.

The second question concerns materiality. Tang and Cooper (2025, p. 731) argue that the epistemic authority of science stems from its connection to the physical world, whereas a system generating text from statistical regularities in a corpus is "devoid of any connection to the physical world". Such a tool may describe an experiment fluently; it has never conducted one.

These two questions intersect at the core of the nature of science, a concept that has long been central to the goals of science education (Erduran & Dagher, 2014; Lederman et al., 2002). If students are expected to understand how scientific knowledge is produced and justified, they must also understand how the output of a generative model is produced and why the two modes of production differ. The empirical foundation for this kind of instruction remains weak. In a systematic review of AI applications in science education, Cheung et al. (2025a) reported that "only three studies mentioned epistemic insights" regarding the relation between scientific knowledge and AI-generated knowledge. Hence, it can be said that the intersection between the nature of science and the epistemology of GenAI remains a genuine gap in the literature rather than an established area of discussion.

The contribution claimed here should be stated against the work that already occupies the neighboring ground. Tang and Cooper (2025) establish materiality as an objection. Cheung et al. (2025a) map the intersection and call for systematic comparison. Rivera-Novoa and Duarte Arias (2025) analyze generative systems as cognitive artifacts that may substitute for or complement the learner, and Tang (2026) reconceptualizes the model as a dialogic partner in classroom discourse. Work in the epistemic vigilance tradition forms a fourth neighbor: Sperber et al. (2010) characterize vigilance as a set of individual cognitive capacities, and Han and Gutierrez (2026) document its enactment by middle school

students working with generative tools. Each of these accounts characterizes the relation between the learner and the model, or the capacities of the individual learner; none proposes a mechanism linking teacher decisions to collective classroom behavior. The explanatory relationship added here is the following: openness to material feedback is the common organizing dimension of the differences observed across the five categories of the FRA, and this dimension predicts that the verification norm of a classroom behaves as an emergent order with threshold and hysteresis properties. The claim is open to refutation. If the classroom verification norm were found to rise smoothly and proportionally as the teacher-set requirement is raised, and to fall back along the same path when that requirement is relaxed, the distinctive prediction of the framework would fail, and the account would be reduced to the general and already established observation that pedagogical design matters. It should also be noted that adoption and prohibition are two rhetorical poles frequently encountered in public discussion rather than the only available positions; conditional use, discipline-specific restriction, and critical integration are all defended in the literature, and the present framework is a contribution to that middle ground rather than a discovery of it.

This study addresses the gap through three questions. In what ways does the knowledge-producing character of GenAI diverge from institutional science? What implications does this divergence have for the teaching of the nature of science? How can the relation between the student, the generative model, and scientific practice be conceptualized in a way that supports design decisions rather than global endorsement or refusal? In answering the third question, the study draws on synergetic thinking, developed by Haken (2004) and elaborated philosophically by Prigogine and Stengers (1984) and by Morin (2008).

The theoretical background is reviewed first, the analytical approach is then described, and the comparison is presented through the five categories of the FRA. A synergetic reframing and a coupled-system model follow, and implications for science education and a research agenda are derived in the final sections.

THEORETICAL BACKGROUND

The Nature of Science: From Consensus Lists to the Family Resemblance Approach

The nature of science has occupied science education research for decades, both as a curriculum goal and as a research framework. In its most common form it has been operationalized as a list of consensus principles: scientific knowledge is empirically grounded, tentative, theory-laden, partly a product of human creativity, and socially and culturally embedded, and there is no single scientific method (Lederman et al., 2002). The consensus

view has proven productive for assessment and teacher education, and the views of nature of science family has become the standard tool in this line of research (Abd-El-Khalick et al., 2024). Criticisms have nonetheless been raised that it presents science as a static set of declarative statements; van Dijk (2011) argued that consensus lists do not reflect the actual diversity of the sciences. Recent evidence indicates that the difficulties the consensus view was meant to address persist in practice. Misconceptions about the tentative character of scientific knowledge and about the relation between laws and theories remain common among students and, to a lesser degree, among teachers (Gyamfi & Ayaaba, 2026), and science was defined as accumulated knowledge even by pre-service teachers who expressed informed views on individual aspects of the nature of science (Sengul, 2023).

The FRA was developed in response to these criticisms. Irzik and Nola (2011) argued that the sciences do not share a fixed essence but, like family members, overlap in some characteristics and diverge in others. Erduran and Dagher (2014) reconceptualized this idea for science education. In that framework, science is a cognitive-epistemic and social-institutional system encompassing the categories of aims and values, methods and methodological rules, practices, knowledge, and social-institutional structures. Kaya and Erduran (2016) distinguished the educational adaptation as the reconceptualized family resemblance approach (FRN). Hence, it is not surprising to see that the number of empirical studies based on the approach has grown rapidly (Erduran et al., 2019), and that a systematic review of the approach itself has been conducted (Cheung & Erduran, 2023).

The relation between the present analysis and the expanded FRN wheel requires a word of explanation, since a simplification has been made. The cognitive-epistemic categories are carried over intact. On the social-institutional side, the FRN account distinguishes several dimensions, among them professional activities, social certification and dissemination, social values, and the financial and political systems within which science operates. These are collapsed here into a single comparative row. What is retained is the mechanism by which claims are subjected to collective scrutiny, namely peer review, the allocation of credit, disciplinary norms, and public accountability, since that mechanism is what the comparison turns on. What is lost is the finer analysis of the political economy of knowledge production, and the loss is not trivial for the second side of the comparison; the funding structures, ownership arrangements, and governance decisions surrounding generative systems are themselves an object of legitimate epistemological attention (Avraamidou, 2024; Olaniyan et al., 2026). This limitation is revisited later.

The suitability of FRA for this study stems from two characteristics. The approach treats science as a system rather than a list and, through the family metaphor,

offers a structure open to comparison; whether a particular knowledge-producing activity belongs to the family is a natural extension of the framework. Cheung et al. (2025a) used the same framework in their review of AI in science education and highlighted the need for a systematic comparison of the epistemologies of science and AI, and the present study responds to that call. A recent critique of the framework itself is also worth noting. Xu and Fan (2024, p. 104) argued that both the consensus view and FRA are “limited by their adherence to staticism and universalism”, calling for a dynamic systems perspective on the understanding of the nature of science. This critique provides the internal justification for the synergetic reframing developed in the sections that follow.

The Epistemic Character of Generative Artificial Intelligence

Large language models generate text by predicting the likely continuation of a sequence from statistical regularities learned from a large training corpus. Bender et al. (2021), in a position paper, characterized these systems as “stochastic parrots” and emphasized that fluent output does not imply commitment to meaning or accuracy. Although the specifics of that characterization have been debated, the point about training remains valid: the objective function rewards probable continuations, while accuracy enters only indirectly, through the composition of the dataset and through alignment based on human preference judgements. Steyvers et al. (2025) examined a measurable consequence of this arrangement. In their experiments, the confidence users attributed to model answers systematically exceeded the models’ internal confidence, and longer explanations increased user confidence even where they did not improve accuracy. The same experiments, however, showed that adjusting explanations to reflect internal confidence narrowed both the calibration gap and the discrimination gap. The problem of uncertainty communication is therefore conditional upon design and configuration rather than an invariant property of the technology, and a survey of uncertainty quantification methods supports the same reading (Liu et al., 2025).

A second line of inquiry addresses the opacity of these systems. Yampolskiy (2020) argued that the inner workings of large models cannot in principle be explained at the level of individual outputs, and Zednik (2021) specified what different stakeholders need to know when dealing with opaque computational systems. Heersmink et al. (2024, p. 1) argue that the coexistence of algorithmic opacity with phenomenological transparency makes it “difficult for users to calibrate their trust correctly”, a point Alvarado (2023) reaches from the character of AI as an epistemic technology. Trust calibration is revisited below as the classroom-level counterpart of epistemic vigilance.

Bearman and Ajjawi (2023, p. 1160) ground a practical conclusion in the assertion that “by definition, AI must always act as a ‘black box’”. Pedagogy, on their account, should address how to work with the black box rather than how to see inside it. This position is adopted here with an extension: working with the black box in science education also requires an explicit account of how the black box differs epistemically from science.

A third line of research examines the contribution of these systems to science itself (Binz et al., 2025; Birhane et al., 2023; Krenn et al., 2022). Generally speaking, the picture emerging from these discussions is that generative systems already contribute to scientific work, and a classroom message declaring such tools to be non-scientific is therefore difficult to defend. It is believed that the coexistence of participation with a fundamentally different style of knowledge production is what makes the situation pedagogically demanding. Hila (2026) has articulated the same tension in the language of justification theory, arguing that large language models can convey knowledge in a reliable sense while lacking reflective justification for the content they convey.

Counter-Accounts and Their Limits

Three counter-accounts stand against the contrast developed in this article, and each is stated here in its strongest form before its limits are considered. The first concerns grounding. Autonomous experimentation platforms, often described as self-driving laboratories, combine machine learning with robotic instrumentation to design experiments, execute them, analyze the results, and revise the next round of hypotheses (Tom et al., 2024), and cloud-based arrangements extend such access beyond the institutions owning the hardware (Canty et al., 2025). In these systems an AI component does close the loop with material feedback, so a claim that AI as such cannot make evidential contact with the world would be false. The second concerns uncertainty: calibration is improvable through explanation design and dedicated estimation procedures, so miscalibrated assertiveness describes the current default configuration rather than a necessary property. The third concerns participation in science, which the work cited above establishes.

A distinction must therefore be drawn between the limitations of particular configurations and the properties of an entire technology family. The first counter-account is decisive against the broader claim and is the reason the scope of this article was restricted: the comparison concerns text-based conversational systems as they are consulted by learners, not AI in general, and not arrangements in which an AI component is embedded within an instrumented experimental loop. The second narrows rather than defeats the claim, and the comparison states the

calibration property conditionally for that reason. The third is not met by restricting the scope, since the systems taking part in scientific workflows are the same conversational systems learners consult; it is met by the argument, where such contributions are shown to remain subject to an exchange of evidence that the models themselves do not perform. What the counter-accounts do not touch is the position of the learner's interlocutor, which performs no measurement and whose output is not revised by the material world between one turn and the next.

The Materiality Objection

The philosophy of science has long emphasized that scientific knowledge is justified through intervention in the material world as well as through representation. Hacking (1983) placed that distinction at the center of his account; confidence in the reality of theoretical entities arises from their functioning within an experimental setup, so that when something can be done with an entity, the ground for speaking about it changes. Pickering (1995) described scientific inquiry as an endeavor in which material agency resists human intentions and knowledge emerges from mutual adaptation. The implication of both accounts for the present study is that justification requires a relation that allows the world to overturn expectations.

Science education research has produced its own body of work along these lines. Wickman (2004) analyzed learning in laboratory work through the concept of practical epistemologies, Hetherington et al. (2018) proposed a material-dialogic theoretical framework, and Tang (2022) showed that material inquiry and material transformation are prerequisites for scientific argumentation. The value of laboratory work for learning outcomes has also been documented through systematic reviews (Agustian et al., 2022; Gericke et al., 2023).

Tang and Cooper (2025) extended this tradition to the debate on GenAI. On their account, a language model operates on statistical probabilities derived from text and has no connection to the physical world, and the fluency of its outputs masks that absence. The implications of this objection for science education are clear. School science already struggles to convey that laboratory work is not merely a matter of verifying textbook statements (Wickman, 2004). When students can obtain a fluent narrative of any experiment at any time, the tendency to equate the narrative with the experiment is likely to strengthen. The synergetic framework developed below positions the materiality objection as a boundary condition rather than as a conclusion; materiality is the link that keeps the wider learning system open to reality.

ANALYTICAL APPROACH

Design and Analytical Steps

A conceptual design was adopted in this study. The design follows the theory adaptation approach described by Jaakkola (2020), in which an established framework is transferred to a new domain in order to transform the way a problem area is understood. Three analytical steps were carried out.

In the first step, the concepts of epistemic authority, opacity, materiality, and epistemic vigilance were clarified through conceptual analysis; epistemic vigilance is understood here, in line with Sperber et al. (2010), as a set of capacities that enable recipients to calibrate their trust in communicated information. In the second step, the five categories of the FRA were used as a comparative analytical framework, and institutional science and GenAI were examined side by side within each category. In the third step, the comparison was reinterpreted through the synergetic concepts of Haken (2004), Prigogine and Stengers (1984), and Morin (2008), and a coupled-system model of the science classroom was constructed. The ordering of these steps is deliberate. The comparison identifies where the differences lie, while the synergetic reframing proposes why the differences take this form and how they may be addressed pedagogically.

The contribution of a conceptual study rests on the traceability of its reasoning rather than on the representativeness of a sample. The remainder of this section therefore reports how the literature was identified, how sources of differing evidentiary status were weighted, how conflicting claims and counterexamples were handled, and how a single source is connected to an analytical category and, in turn, to a testable proposition. A systematic review protocol is neither claimed nor appropriate here; what is presented is a transparent conceptual procedure.

Identification, Selection, and Weighting of the Literature

Search scope

The literature was identified through structured searches in Scopus, the Web of Science core collection, and ERIC. These searches were complemented by backward and forward citation chaining from the framework-defining sources (Erduran & Dagher, 2014; Irzik & Nola, 2011; Tang & Cooper, 2025). Three families of terms were combined: nature-of-science terms such as nature of science, FRA, and epistemic cognition; GenAI terms such as GenAI, large language model, and ChatGPT; and epistemic terms such as epistemic authority, epistemic vigilance, materiality, calibration, and opacity.

Table 1. Literature clusters, selection criteria, and evidentiary role in the analysis

Cluster	Included	Evidentiary role	Weighting rule
Nature of science	Framework-defining works, systematic reviews, empirical studies employing the FRA	Supplies the comparative categories and the educational aims against which divergence is assessed	Framework-defining works are treated as definitional rather than evidential
Epistemic character of generative artificial intelligence	Peer-reviewed philosophical and ethical analyses; technical studies reporting measured properties	Supplies the characterization of the second side of the comparison	Measured properties such as calibration and fabrication outrank characterizations
Generative artificial intelligence in science education	Systematic reviews and meta-syntheses; individual empirical studies for within-category tensions	Supplies classroom-level consequences and the warrant for the propositions	Review evidence outranks single studies; opinion pieces are cited as arguments
Synergetic and complexity	Primary works in synergetic and in dissipative-structure theory; complexity-in-education sources	Supplies the reframing vocabulary and the constraints upon it as an analogy	Used analogically; no empirical claim is transferred from physics

Foundational works in the philosophy of science and in synergetic were included irrespective of date. For the GenAI literature the window was November 2022 to September 2026, the opening of which corresponds to the public release of conversational large language models.

Selection

Within the nature-of-science cluster, framework-defining works, systematic reviews, and empirical studies employing the FRA were included. Within the GenAI cluster, peer-reviewed philosophical and ethical analyses were included together with technical studies reporting measured properties such as calibration or fabrication rates. At the intersection, systematic reviews and meta-syntheses were prioritized, and individual empirical studies were consulted where a within-category tension required illustration. Studies addressing implementation, attitudes, or acceptance without epistemic content were excluded, as were studies of non-generative educational technologies.

Evidentiary status

Each source was coded for the kind of warrant it can carry philosophical argument, opinion or position paper, review evidence, and primary empirical evidence. This coding is reflected in the text. Arguments are described as arguments, and empirical claims are attributed to the studies that produced them rather than to the commentaries citing them. Preprints were admitted in an illustrative or corroborating role only, and they were not used as the sole warrant for a comparative judgement. Where a preprint supplied a claim carrying analytical weight, a peer-reviewed source is cited alongside it.

Table 1 presents the literature clusters, the criteria by which sources were included, and the evidentiary role assigned to each cluster in the analysis.

Conflicting claims and counterexamples

Where sources disagreed, priority was given to primary empirical evidence over review evidence, and to review evidence over argument. Where disagreement persisted at the same evidentiary level, the claim was reformulated conditionally, and its boundary was stated in the text or in the corresponding row of the comparison presented later. Counterexamples were not set aside. Each counterexample was assessed for whether it narrows the scope of a claim, renders the claim conditional upon a configuration, or defeats the claim altogether. Instances of the first two kinds are recorded as boundary conditions in the comparison presented later. The most consequential instance is that of autonomous experimentation platforms, in which an AI system does close the loop with material feedback (Canty et al., 2025; Tom et al., 2024). Hence, it can be said that the argument developed here holds for text-based conversational systems as they are encountered by learners, rather than for AI as such.

A Worked Example of the Analytical Chain

Two chains are traced from source to proposition in order to make the procedure inspectable. The chains are reconstructions of decisions already taken rather than derivations performed for the first time here; what they are intended to show is that each comparative judgement can be followed back to a source and to an interpretive decision that could have gone the other way.

The first chain concerns the category of scientific knowledge. Steyvers et al. (2025) report a calibration gap between users' confidence in model answers and the models' internal confidence, and they also report that this gap narrows when explanations are adjusted to reflect internal confidence. Two readings of the finding are available. Miscalibrated assertiveness may be an invariant property of the technology, or it may be a designable feature of a given configuration. The second reading was adopted, since the same study reports the manipulation that narrows the gap, and survey evidence

Table 2. Institutional science and GenAI compared across the family resemblance categories

FRA category	Epistemology of institutional science	Operation of a text-based generative model	Boundary conditions and counterexamples	Tension in the science classroom
Aims and values	Approximation to truth, explanatory coherence, and predictive accuracy, guided by empirical adequacy and organized skepticism	Maximization of sequence likelihood conditioned on training data and alignment preferences; fluency rather than truth is the operative target	Alignment procedures can raise factual accuracy as a proximate objective; the values remain those of the developer rather than of an epistemic community	Students may read linguistic fluency as an indicator of epistemic reliability
Methods	Experimentation, systematic observation, modelling, and the elimination of error through material feedback	Pattern completion over a corpus; no capacity to intervene in the physical world and no new evidential contact at the time of inference	Retrieval and tool use extend the set of accessible records (Lewis et al., 2020) but generate no new measurement; autonomous experimentation platforms do close the loop and fall outside the scope of this comparison	Consulting the model may displace firsthand investigation
Practices	Argumentation based on evidence, material inquiry and transformation, and peer critique	Simulation of argumentative discourse; the evidence invoked has been examined by others rather than by the system	Retrieval-grounded output can cite records, and the citations can be checked; the checking is performed by the reader	Argumentation practice risks becoming rhetorical rather than evidential
Knowledge	Revisable networks of theories, laws, and models calibrated against observation	Parametric associations without propositional commitment; outputs may contain fabricated content, and assertiveness is weakly calibrated to accuracy	Calibration is configuration-dependent and improvable through explanation design and uncertainty estimation (Liu et al., 2025; Steyvers et al., 2025) expressed output	The tentative character of scientific knowledge contrasts with confidently expressed output
Social-institutional system	Peer review, allocation of credit, disciplinary norms, and public accountability	Opaque training corpora and corporate governance; alignment procedures exist but do not constitute epistemic peer evaluation	Accountability in science is a normative achievement that can fail, as replication difficulties and retractions show; the contrast is one of mechanism rather than of guaranteed outcome	Epistemic trust must be taught as a spectrum rather than as a binary

on uncertainty quantification supports treating calibration as configuration-dependent rather than fixed (Liu et al., 2025). The interpretive consequence is recorded in the knowledge row of **Table 2**, where the calibration property is stated conditionally. The same decision shapes proposition 1, in which fluency and accuracy are manipulated independently instead of being assumed to covary.

The second chain concerns the category of methods. Lewis et al. (2020) establish that retrieval augmentation extends the set of texts available at inference, while the materiality criterion of Tang and Cooper (2025) distinguishes access to records from contact with the material world. The interpretive decision was to narrow the closedness claim from the absence of new information at inference to the absence of new evidential contact, so that the claim survives the retrieval counterexample while conceding what that counterexample establishes. The narrowing is stated later, and it is what carries the openness criterion to the level of the system rather than to the individual act.

COMPARATIVE EPISTEMOLOGICAL ANALYSIS THROUGH THE CATEGORIES OF THE FAMILY RESEMBLANCE APPROACH

In this section the comparison is conducted across the five categories of the FRA. **Table 2** summarizes the analysis, and the subsections elaborate on each category in turn. Two qualifications should be stated at the outset. The comparison concerns institutional science as a knowledge system and text-based conversational systems in their currently operational forms, so that it should be read as time-dependent rather than essentialist. The two sides are also not ontologically equivalent, since science is a social-epistemic institution whereas a generative model is an artifact. The encounter in the classroom nonetheless takes place at a functional level; the student consults the model as a source of knowledge, as a textbook or a teacher would be consulted. The comparison therefore concerns the epistemic character of two sources that fulfil the same function. Boundary conditions and counterexamples are

recorded in the final column of **Table 2** and are discussed before.

Aims and Values

Science pursues explanatory coherence, predictive accuracy, and successive approximations to truth under shared values such as empirical adequacy, honesty in reporting, and organized skepticism (Kuhn, 1970; Longino, 1990; Merton, 1973). These values are socially enforced, and they become visible to students when a claim is qualified, cited, or revised. A generative model has no aims of its own. The training objective rewards probable continuations of text (Bender et al., 2021), while alignment procedures priorities the outputs preferred by human evaluators (Ouyang et al., 2022). The operational aims are helpfulness and fluency, and accuracy is served only indirectly, to the extent that the data and the evaluators serve it. The classroom tension arises from this, since fluency is the surface feature students are most likely to read as an indicator of reliability (Heersmink et al., 2024). Cooper (2023) has documented this authority effect in science-specific interactions. It is believed that the effect is stronger among students who equate the image of knowledge with confident answers, and this expectation is formulated as proposition 1 later.

Methods and Methodological Principles

Scientific methods are diverse, yet they share a commitment to the elimination of error through engagement with the world; hypotheses are opened to observation and experiment, and empirical feedback is allowed to overturn expectations (Hacking, 1983; Popper, 1959). A generative model completes patterns over a corpus fixed before the conversation begins (Bender et al., 2021), and at the moment of inference it cannot intervene in the world, collect data, or be surprised by a measurement (Tang & Cooper, 2025). Configurations augmented with search can add online texts to that corpus (Lewis et al., 2020), with consequences discussed later. The methodological virtue displayed in its output derives from the texts of scientists who performed the work. The practical risk in the classroom is displacement, and overreliance on AI dialogue systems has been reported to substitute for analytical effort (Zhai et al., 2024). When an answer can be obtained within seconds, taking the slow path of inquiry requires a justification, and that justification is itself a lesson about the nature of science. Hence, it can be said that the answer and the inquiry are not epistemically equivalent. The relation between the risk of displacement and the mode of model use is addressed in proposition 4 later.

Scientific Practices

Science education reforms increasingly define science through its practices, among which the construction and

interpretation of data and the construction of arguments from evidence are foremost (Erduran & Dagher, 2014; Osborne & Pimentel, 2023). Tang (2022) identified the material prerequisites of classroom argumentation; questions are framed around the behavior of the material world, and evidence is generated through the coordinated arrangement of objects, bodily actions, and words. Generative models imitate the discourse of these practices with considerable skill (Bender et al., 2021; Binz et al., 2025), producing a well-structured argument or a textbook-quality description of a scientific model. What cannot be done is to ground the argument in evidence the system itself has examined (Tang & Cooper, 2025). The imitation is rhetorical rather than evidential, and this holds even where retrieval supplies citations, since the examination of the cited record is performed by the reader. Argumentation instruction has therefore acquired a further task: students must learn to distinguish the form of an argument from its justification and to ask where the evidence stands for every fluent claim (Osborne & Pimentel, 2023; Tang, 2024). The effect of instruction that positions model output as a hypothesis to be tested is examined later through proposition 2.

Scientific Knowledge

Scientific knowledge takes the form of networks of theories, laws, and models that rest on observation, are calibrated against one another, and remain subject to revision (Kuhn, 1970). Tentativeness is a characteristic of the system, which signals its own uncertainty through margins of error, confidence levels, and open disputes. The knowledge embedded in a generative model has a different structure. It consists of parametric relationships distilled from text, carrying no propositional commitment and no internal marking of which relationships are well supported (Bender et al., 2021; Hila, 2026). A false statement and a true statement can be produced with the same apparent confidence, and the weak coupling between linguistic assertiveness and accuracy has been measured directly (Steyvers et al., 2025). That coupling is nonetheless improvable; what is described is the default configuration rather than a necessary property. Fabricated content is not an occasional accident. In a classroom verification exercise in which students checked 456 model-generated references against the literature, 51.3% of those references proved to be real and correctly cited, and the fabricated ones clustered in disciplinary journals rather than in high-profile multidisciplinary ones (Sevryugina & Vargas, 2026). For learners, the contrast between the two knowledge structures is instructive when addressed explicitly and misleading when it is not. It is believed that confidently expressed output reinforces the naive view of knowledge as a stock of settled answers, and this is precisely the view that science education seeks to unsettle (Gyamfi & Ayaaba, 2026; Sengul, 2023).

Social-Institutional System

Science operates through institutions that allocate credit, enforce norms, and subject claims to peer review and public accountability (Longino, 1990; Merton, 1973). These structures do not guarantee truth; they create the conditions under which errors can be identified and corrected. The qualification matters, since the mechanism is a normative achievement that can and does fail; difficulties of replication, uneven review quality, and retraction are part of the record, and the contrast drawn here concerns the presence of a correcting mechanism rather than the guarantee of a correct outcome. Generative models are built and governed differently. Training corpora remain largely undisclosed (Bender et al., 2021), design decisions are made within companies (Avraamidou, 2024), and no mechanism comparable to peer review examines individual outputs (Birhane et al., 2023). The educational implication concerns epistemic trust. Students already navigate a media environment in which sources vary sharply in reliability (Osborne & Pimentel, 2023), and the generative model adds a source with opaque institutional backing. Jose et al. (2025, p. 2), in an opinion article, describe the classroom consequence of this integration as GenAI becoming “a new actor in the epistemic space”, and they argue that authority may shift quietly from human educators to algorithms. Empirical work is consistent with the concern without yet settling it; a survey of higher education students found extensive use of AI assistants alongside reliance on inadequate verification methods (Martin-Moncunill & Alonso Martinez, 2025). Hence, it can be said that teaching differentiated trust has become part of the social-institutional dimension of science education, rather than a matter of acceptance or rejection.

THE SYNERGETIC REFRAMING OF THE RELATION BETWEEN STUDENTS, THE GENERATIVE MODEL, AND SCIENCE

Synergetic Foundations

The comparative analysis shows that text-based conversational systems diverge from science in every category of family resemblance. What the analysis has not yet supplied is a unifying explanation of why the differences follow this pattern, nor a constructive model for how the classroom should be organized around them. Synergetics offers both. Developed by Haken (2004) as a theory of self-organization, synergetic examines how open systems far from equilibrium produce ordered structures. Three concepts are drawn upon here. The order parameter is a macroscopic variable that emerges from the interactions of system components and then constrains them, a relation Haken (2004) terms the slaving principle. The control parameter can be adjusted from outside and can drive the system

towards qualitative transitions. A phase transition is the point at which one mode of organization gives way to another. Prigogine and Stengers (1984) showed that dissipative structures of this kind depend on continuous exchange with the environment, while Morin (2008) developed a corresponding epistemology in which knowledge is an emergent property of relationships rather than a property of parts. The extension of this tradition to AI has already been attempted. Muminova (2025) reads GenAI as a nonlinear, self-organizing phenomenon shaped by recursive interactions among algorithms, data ecologies, and human agency, an argument that supports the transfer of synergetic concepts to the present problem without settling how those concepts should be operationalized in a classroom. In cognitive science, Favela and Amon (2023, p. 1) have advocated raising complexity, which includes synergetic, to the status of “complexity science as a centerpiece of the discipline”; a similar call has not yet been taken up in science education.

Open Systems and Closed Systems

From a synergetic perspective, the difference between science and a generative model can be expressed through a single dichotomy. Science is an open system. It maintains its ordered structure, that is, the coherent body of scientific knowledge, through a continuous exchange of evidence with material reality, and this order is dissipative in the sense defined by Prigogine; if the exchange were to cease, the structure would be expected to degenerate into dogma (Prigogine & Stengers, 1984). A generative model is operationally closed with respect to evidence at the moment of inference. It rearranges patterns within a corpus fixed in advance, and however refined the rearrangement, no new evidential contact with the world takes place. The order observed in its output is therefore parasitic upon the openness of science, and the term is used in the sense of derived dependence rather than as a metaphor of harm. This formulation incorporates the materiality objection of Tang and Cooper (2025) into a systems framework and accounts for the pattern observed in the comparative analysis. All five categories, namely aims, methods, practices, knowledge, and institutions, point in the same direction, since each is organized in part by the same feature: openness to material feedback, which science possesses and the conversational model lacks. The categories retain their own content, and the claim is that this feature accounts for the direction in which they differ rather than for everything they contain. In the first four categories the connection is direct. In the social-institutional category the connection is indirect but structural, since peer review, the allocation of credit, and accountability can be interpreted as mechanisms that socially regulate the link between claims and evidence (Longino, 1990). On this interpretation, institutions are structures that institutionalize the exchange of evidence,

and a system that does not itself engage in that exchange has developed no corresponding institutions.

A current objection to this contrast is that some architectures are equipped with search and tool use, so that a network-connected model can access new texts at inference. The objection delineates the limits of the closedness claim without invalidating it. Search and tool use expand the set of texts the model can reach; they do not alter the nature of evidential contact, since the model still operates on records produced by others and cannot construct an experimental setup, take a measurement, or experience the overturning of an expectation through material feedback. The materiality criterion of Tang and Cooper (2025) draws precisely this distinction. Closedness is therefore a characterization of the type of evidential engagement rather than of the size of the corpus.

A second and more demanding objection concerns the status of the claim itself. If every category of difference is traced to a single property, the analysis risks reintroducing the essentialism that the family resemblance perspective was designed to avoid. Two responses are required. The first concerns the logical form of the claim. Openness to material feedback is proposed here as the organizing dimension that accounts for the observed pattern of differences, not as a necessary and sufficient criterion for membership in the family of sciences. It is a family-level tendency rather than a defining essence, which is the distinction the family resemblance perspective was introduced to preserve (Irzik & Nola, 2011), and the categories retain their own content. The second response concerns the boundary cases, which are instructive rather than inconvenient. Observational sciences such as astronomy, paleontology, and epidemiology make no intervention in Hacking's (1983) sense, yet they remain in continuous evidential contact with their objects. Theoretical and mathematical work within the sciences produces no measurements at all and derives its accountability from the wider system in which it is embedded. Computational and simulation-based science operates on models whose calibration against measurement is what secures their standing. Large distributed collaborations place most participants at several removes from the instrument. It is not possible to make a generalization about the openness criterion at the level of the individual researcher, and the criterion is therefore defined at the level of the system: what matters is that the body of knowledge as a whole remains answerable to evidence that the world can overturn. The same system-level reading applies to the other side of the comparison. A generative model is also embedded in a sociotechnical system of developers, benchmarks, and human feedback; what distinguishes that system is that its corrective loop closes on human preference judgements rather than on material resistance.

This system-level formulation carries a consequence for the learner. A clarification is required concerning what the openness criterion demands of learners and of scientists. Direct experimentation is not the only epistemically productive route to evidence. Researchers and students rely legitimately and routinely on evidence produced by others, including archived observations, instrument outputs, published datasets, and expert testimony. Mediated reliance of this kind becomes epistemically productive through four features of its structure. The origin of a claim can be traced to a procedure and to a source; the procedure carries an account of its own error; someone is answerable for the claim; and the claim can be contested through channels that could in principle overturn it. Recent analyses of epistemic dependence in learning mediated by AI converge on a similar set of criteria for separating productive reliance from harmful dependence (Du & Yuan, 2026), and epistemological treatments of artificial testimony argue that machine-generated statements can ground belief only where the testifier reliably exhibits properties analogous to veracity and cautiousness (He & Yang, 2025).

The weakness of generative mediation is specific rather than global, and it is located in these four features. An individual output does not carry provenance for the claim it contains, even where retrieval supplies citations for the passage as a whole. It carries no error model, so that its linguistic assertiveness cannot be read as a rate of being correct. No agent is answerable for it, and the systems producing it have been characterized as lacking the reflective regulation and normative commitment that answerability presupposes (Woodruff & Hewitt, 2026). Contestation of a single output has no channel through which the source itself could be revised. Hence, it can be said that the openness criterion is a property of the classroom system as a whole rather than of each individual act within it, and that students are not required to generate every measurement themselves.

The Coupled-System Model for Science Classrooms

The pedagogical question is how to organize a classroom in which an open system, a closed system, and a learner interact. **Figure 1** presents the proposed model. The student, the generative model, and scientific practice are treated as coupled subsystems. The student engages in dialogic exchange with the model and in inquiry and evidence-based practices with science, while the model remains dependent on the scientific record from which its corpus derives.

The arrangement is organized by two synergetic structures. These two structures are kept analytically distinct, since conflating them would make the model explain verification through itself.

The first structure is the teacher-set verification requirement, which functions as a control parameter. It

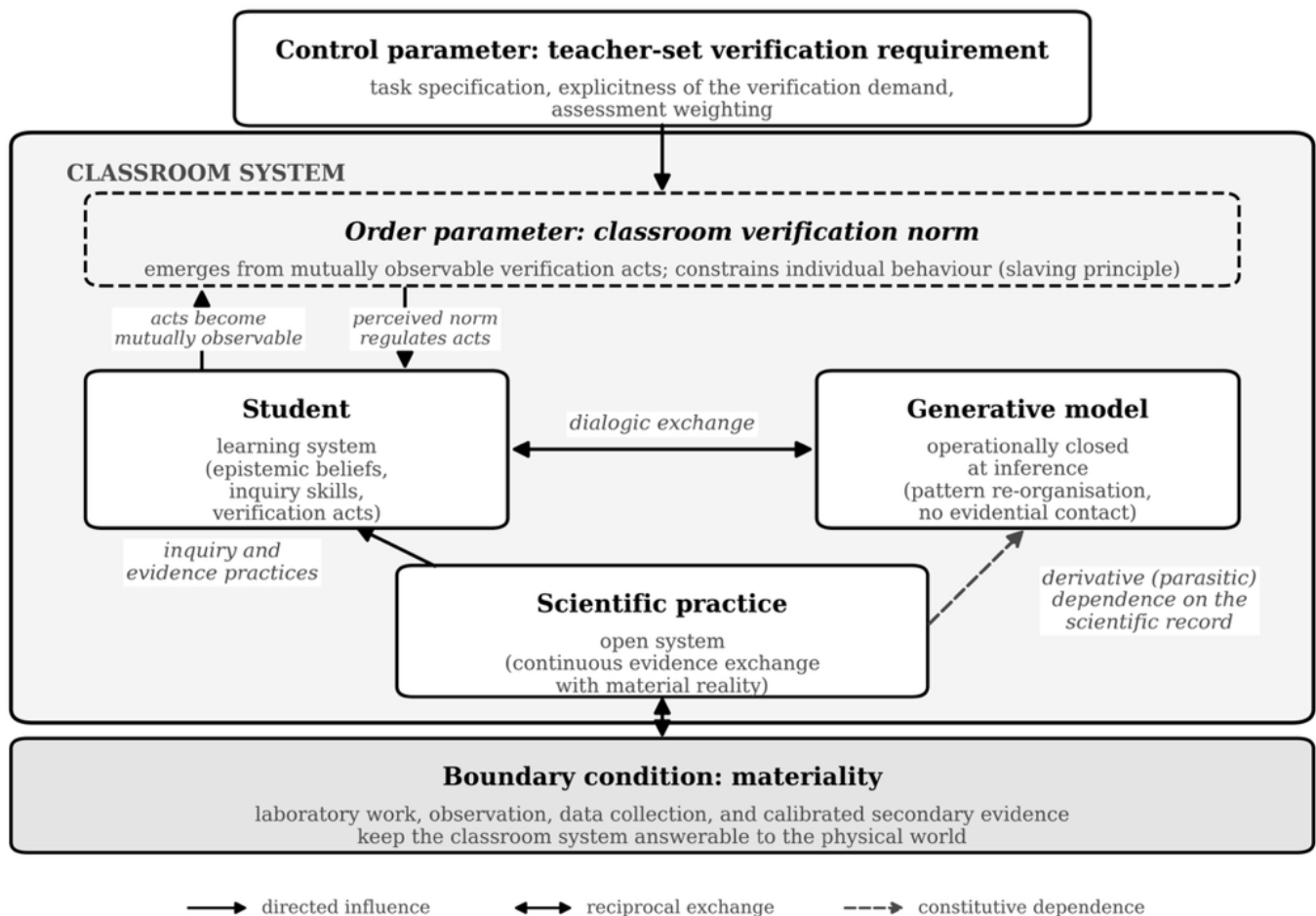


Figure 1. A synergetic coupled-system model of the student, the generative model, and scientific practice (created by the authors)

is exogenous and adjustable in graded steps: how explicitly a task states that model output is to be checked against an independent source, how much of the assessed work depends on that check, and how clearly an adequate check is defined. The requirement is a property of the designed environment, readable from task specifications and assessment rubrics, and it does not presuppose that any student verifies anything.

The second structure is the classroom verification norm, which is proposed as the order parameter. It is endogenous, and it emerges from the aggregate of students' verification acts as those acts become mutually observable. The norm is a property of the collective rather than of the task, and it is measurable through the frequency of verification acts in classroom discourse together with students' perception of what is ordinarily done in their classroom. Once established, the norm constrains individual behavior in the manner Haken (2004) describes through the slaving principle: students scrutinize claims because that is what is done in this classroom, and not merely because a task requires it.

The distinction is temporal as well. The requirement is set at the outset and varied by the teacher, whereas the norm forms later, and its formation is the phenomenon to be explained. The mechanism runs in a loop of four

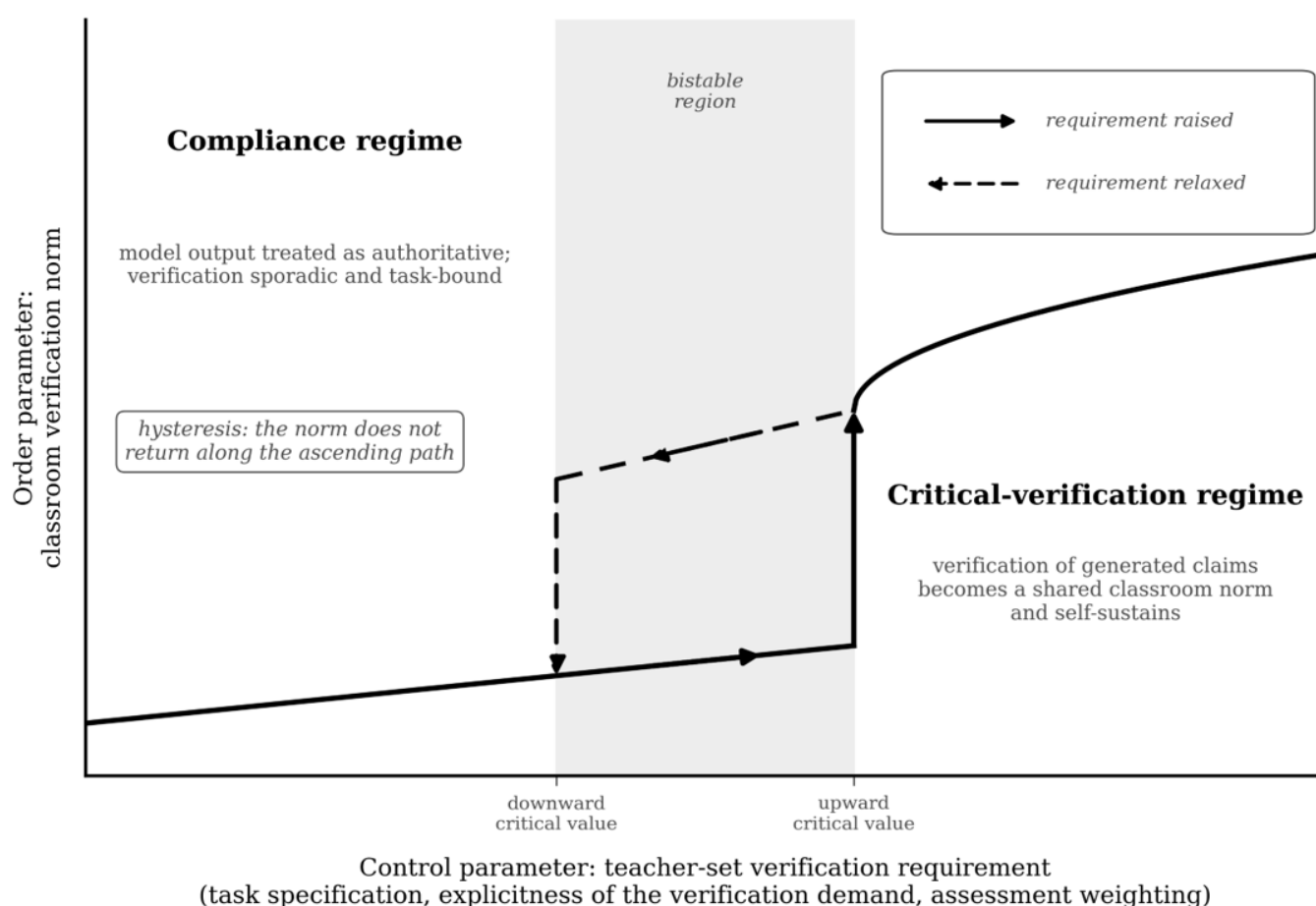
steps. A requirement prompts individual verification acts; the acts become visible in classroom discourse and are taken up by peers; their visibility shifts what students perceive to be the ordinary practice of the class; and the perceived practice then regulates individual acts, including those of students for whom the requirement was never binding. This fourth step distinguishes an emergent norm from compliance, and it is what makes the norm capable of outlasting the requirement that seeded it. **Table 3** presents the two parameters with their indicators and timing.

The synergetic prediction is that the effect of the control parameter is not linear. Below the critical value, the class is expected to settle into a regime in which model output is treated as authoritative; above it, a regime of critical verification is expected to self-organize and to sustain itself. **Figure 2** illustrates this threshold behavior together with the path followed when the requirement is subsequently relaxed.

Materiality enters the coupled-system model as a boundary condition. Laboratory work, observation, and data collection are the links that keep the system as a whole open, and it is for this reason that the proposed model treats evidential accountability to the material

Table 3. The control parameter and the order parameter: Definitions, indicators, and timing

	Teacher-set verification requirement (control parameter)	Classroom verification norm (order parameter)
Locus	The designed environment	Collective behavior
Status	Exogenous, set by the teacher	Endogenous, emergent from interaction
Indicators	Task specification, explicitness of the verification demand, assessment weighting	Observed frequency of verification acts; perceived descriptive norm; density of verification moves in classroom discourse
Timing	Set at the outset and varied in graded steps	Measured repeatedly thereafter; formation is lagged
Role in the model	Drives the system towards or away from the transition	The ordered state whose onset and persistence are to be explained

**Figure 2.** Schematic representation of the transition between classroom regimes as a function of the teacher-set verification requirement, with the hysteresis loop followed on partial withdrawal (created by the authors)

world as structurally necessary rather than merely desirable.

Three clarifications are necessary to prevent the coupled-system model from being overgeneralized. The term “synergetic” is used here as a structured analogy with specific rules of correspondence, in the manner proposed for concepts of complexity in education (Koopmans, 2020); no claim is made that classroom dynamics conform to the equations of laser physics. The coupled-system model further does not deny that generative systems contribute to contemporary science (Binz et al., 2025; Krenn et al., 2022). It specifies the nature of that contribution: the models contribute within scientific practice, and their outputs are subject to the

exchange of evidence that the models themselves cannot perform.

The third clarification concerns what the synergetic vocabulary adds. That verification behavior is shaped by shared classroom expectations can be stated without any reference to synergetic, and three existing accounts state it well. Research on classroom norms explains how expectations about legitimate practice regulate participation, and how such norms are negotiated rather than imposed (Yackel & Cobb, 1996). Work on metacognitive scaffolding has shown that structured prompts raise students’ questioning and verification during collaboration supported by GenAI (Zhuo & Bai, 2026). The Apt-AIR framework accounts for how

learners come to evaluate knowledge claims aptly and how education can support that capacity (Barzilai & Chinn, 2024). A recent systematic review reaches a conclusion consistent with all three, reporting that epistemic outcomes are governed by pedagogical design, AI literacy, and assessment practices rather than by the technology itself (Suazo Galdames et al., 2026).

The framework proposed here does not compete with these accounts over whether design matters. Its claim concerns the shape of the relation between design and collective behavior, and four predictions follow from the synergetic reading that do not follow from a norm-based, a scaffolding-based, or an epistemic-cognition account. As the verification requirement is raised in graded steps, the classroom norm is expected to change discontinuously at a critical value rather than proportionally. Once the transition has occurred, partial withdrawal of the requirement is not expected to return the norm to its prior level, so that the system exhibits hysteresis. Variability and temporal autocorrelation in verification behavior are expected to increase as the critical value is approached, as they do near transitions in other self-organizing systems. The distribution of classrooms held at intermediate requirement levels is expected to be bimodal rather than centered, since intermediate collective states are unstable. A framework predicting only that verification improves under better design would add nothing to the accounts above. Two of these predictions, the discontinuity and the hysteresis, constitute proposition 5 later; the remaining two are stated there as auxiliary indicators of the same transition. Their failure would count against the framework rather than against the general claim that classroom design matters.

IMPLICATIONS FOR SCIENCE EDUCATION

Evidential Accountability and the Place of Practical Work

The coupled-system model provides a renewed rationale for practical work. If openness to material feedback is the boundary condition that keeps the learning system from closing, then laboratory and field activities serve as a foundation rather than as enrichment. An effective teaching pattern follows: generative output is positioned as a hypothesis and the material world as the arbiter, so that students obtain a prediction from the model, conduct the investigation, and account for any discrepancy. Activities of this kind turn the distinction between closed and open systems into a lived one, and they answer the concern that a fluent description may be conflated with empirical contact (Tang & Cooper, 2025). Research on the learning outcomes of laboratory work supplies the empirical foundation of the pattern (Agustian et al., 2022; Gericke et al., 2023).

Two qualifications keep this recommendation within defensible limits. The first concerns the channel. What the framework requires is evidential accountability to the material world, and firsthand practical work is the most direct channel to it rather than the only one; observational datasets, remotely operated instruments, citizen science archives, and simulations calibrated against measurement can sustain the same accountability where laboratory access is limited. A blanket claim that practical work is necessary in every learning arrangement would not survive contact with the observational sciences discussed before. The second qualification concerns the weighting of evidence. A measurement a student has taken personally does not automatically outrank published evidence of higher quality, and measurement quality, sample size, and the conditions of collection belong in the classroom conversation alongside the provenance of the claim.

Table 4 presents an illustrative sequence built on this pattern. The example is offered as an illustration of the design logic rather than as a validated instructional unit.

Argumentation and the Evidential Status of Generative Output

Argumentation instruction acquires a further component: the placing of model output within an epistemic hierarchy. Students can be taught to treat a generated claim as testimony from an opaque source and to weigh it against data they have collected themselves, published sources, and expert consensus. Tang (2024) has proposed epistemic metacognitive questions for AI-generated texts, foremost among them how the evidence can be verified and which explanations have been left out. Classroom protocols that search for the evidence implicitly assumed in a coherent paragraph activate the vigilance the coupled-system model defines as the order parameter. Recent evidence indicates that the activation is achievable with explicit support: epistemic criteria and question prompts together supported middle school students in appraising explanation quality, checking source credibility, and reflecting on their own assumptions while working with generative tools (Han & Gutierrez, 2026). Sperber et al. (2010) noted that vigilance is socially calibrated, and the classroom is the setting in which that calibration can be shaped deliberately.

Assessment and Evaluative Judgement

Assessment regimes are among the most powerful control parameters available to the teacher. Regimes rewarding the polished product invite transfer of the task to the model, whereas regimes measuring the inquiry process, the documentation of verification, and the quality of evaluative judgement reward behaviors the model cannot perform. Gao et al. (2025) reclassified science assessment along the dimensions of

Table 4. An illustrative instructional sequence positioning model output as a hypothesis

Element	Illustration
Level and topic	Lower-secondary physical science; the variables affecting the period of a simple pendulum
Phase 1: elicitation	Students ask the model which variables affect the period and why, and record the answer together with the justification offered for it
Phase 2: commitment	Before any measurement, each group records a written prediction and states what observation would count against it
Phase 3: investigation	Length, mass, and amplitude are varied systematically; repeated timings are taken and the spread across trials is recorded
Phase 4: confrontation	The measured pattern is compared with the model's answer, and discrepancies are classified as measurement error, an unstated boundary condition, or an incorrect claim
Phase 5: provenance	Groups locate a published value or dataset, compare it with their own result, and decide which evidence is stronger and why
Teacher prompts	What evidence would the model have needed in order to know this? Where did our result and its answer diverge, and which of the two can be checked? What would make our measurement more trustworthy?
Assessment criteria	The link between claim and evidence; the statement of provenance for each source; the account given of the discrepancy; the calibration of expressed confidence to the evidence held

multimodality and interactivity and showed that tasks high on both dimensions resist delegation to AI, which is the concrete expression of the control parameter framework in assessment. Bearman and Ajjawi (2023) argued that working with opaque systems requires judgement about when and how far to trust them, while Zhai et al. (2024) reported that unchecked trust can erode the competencies education seeks to develop. Hence, it is not surprising to see that early empirical studies report both productive and corrosive patterns of use; the difference is thought to lie in the surrounding design, and a recent systematic review concludes that epistemic outcomes are governed by pedagogical design, AI literacy, and assessment practices rather than by the technology itself (Suazo Galdames et al., 2026). The language of control parameters makes that design explicit.

Teacher Education

Control parameters are in the hands of the teacher, and teacher preparation therefore carries considerable weight in the proposed model. Garofalo and Farenga (2025) found that science teachers at the outset of the generative era expressed cautious optimism while voicing concerns about knowledge production and original thought; although expressed in practical terms, these concerns are epistemological. Jose et al. (2025) argue that teacher authority may be subjected to algorithmic comparison, so that teachers may find themselves defending their own reasoning against machine output. An instrument for measuring pre-service teachers' perceptions is available (Ishmuradova et al., 2025), these systems have entered science teacher education across economically diverse countries (Lee et al., 2026), and competence frameworks specific to science teaching have appeared (Huwer et al., 2025). The present analysis suggests that teacher education should treat the epistemological comparison between science and GenAI as a distinct component of subject

knowledge, so that teachers can both apply the distinction between open and closed knowledge systems and articulate it.

Privacy, Equitable Access, and Teacher Responsibility

Three practical conditions attach to the classroom applications proposed above.

The first is privacy. Interaction records are evidence for the verification norm and also a record of what individual students did not know; retention periods, access rights, and the level of aggregation require settlement before any systematic collection begins. The second is equitable access. Capable systems are distributed unevenly across devices, subscriptions, and languages, and reliance on them and the freedom to decline them are themselves unequally distributed (Olaniyan et al., 2026). The third is teacher responsibility. The control parameters described here add design and monitoring work to an existing workload, and the framework should not be read as a transfer of institutional responsibility to the individual teacher. It should be noted that the present article is conceptual and involves no participants. The empirical program outlined in the following section, which would involve students and interaction records, requires the corresponding consent and ethical review procedures, and those requirements should not be conflated with the status of this study.

RESEARCH AGENDA AND RECOMMENDATIONS

The framework yields propositions that are open to empirical test. A framework is tested most severely where it makes its strongest claim, and **Table 5** therefore maps each component of the coupled-system model onto the proposition that tests it, together with the level at which the test is located.

Table 5. Mapping between the components of the coupled-system model and the propositions

Model component	Claim	Proposition	Level of analysis
Epistemic authority of fluent output	Perceived credibility tracks fluency rather than accuracy	P1	Individual
Materiality as a boundary condition	Positioning output as a hypothesis to be tested yields greater gains in understanding the nature of science	P2	Individual, within class
Order parameter	The classroom norm constrains individual behavior beyond individual attitudes	P3	Class and individual, multilevel
Mode of engagement, as shifted by the classroom norm	Dependent and autonomous reliance diverge in their relation to inquiry engagement, and the classroom verification norm is what moves students between them	P4	Individual, repeated measures
Control parameter and phase transition	Graded change in the teacher-set requirement produces discontinuous change in the classroom norm, with hysteresis	P5	Class, repeated measures

Proposition 1: The Authority Effect

The perceived credibility of generated output will vary with linguistic fluency independently of factual accuracy, and the effect will be stronger among students holding naive views of the nature of science. Urban et al. (2025) have shown that students' epistemic beliefs and metacognitive accuracy shape how correct and incorrect model-generated content is integrated into their own writing. What remains untested is the independent manipulation of fluency and accuracy with views of the nature of science as a moderator, and this is what the proposition specifies. An experimental vignette design with a fluency-by-accuracy manipulation and a moderation analysis is appropriate.

Proposition 2: Materiality

Instructional sequences that position model output as a hypothesis to be tested empirically will yield greater gains in understanding the nature of science than sequences using the model solely as an explanatory resource, provided that hands-on time is held constant.

Proposition 3: The Order Parameter

Classroom-level verification norms will predict individual verification behavior beyond individual attitudes and prior achievement. Where the distribution of verification acts across students is represented as a node-link structure, classroom epistemic discourse can additionally be modelled through network psychometrics.

Proposition 4: The Mode of Engagement

The relation between model use and inquiry engagement will depend on the mode of reliance rather than on its frequency. Dependent reliance, in which core reasoning is delegated, will be negatively related to inquiry engagement, whereas autonomous reliance, in which the model is used as a scaffold and the reasoning is retained by the student, will not. Evidence from outside science education already points in this direction (Wang & Zhang, 2026; Zhu et al., 2026), and the present

framework predicts that the classroom verification norm is what shifts students between the two modes. Mode and engagement are measured at each wave, and frequency of use alongside them, so that the two predictors can be entered into the same multilevel growth model and the variance explained by mode over and above frequency can be reported. Latent profile analysis is appropriate for recovering configurations of individual use, but it is reported as a secondary and descriptive analysis rather than as evidence of threshold behavior.

Proposition 5: The Control Parameter and the Phase Transition

As the teacher-set verification requirement is raised in graded steps, the classroom verification norm will change discontinuously to a critical value rather than proportionally; and at any given level of the requirement, the norm will be higher when that level is reached on a descending path after the transition than when the same level was passed on the ascending path.

This proposition carries the central claim of the coupled-system model, and it requires a design that the preceding propositions do not. Classes are observed across at least seven waves spanning a baseline phase, three graded increases in the requirement, a plateau, a withdrawal phase, and a follow-up, so that the ascending path is sampled densely enough for a breakpoint to be estimated rather than assumed. The outcome is the classroom verification norm as defined before, measured at each wave. Three competing functional forms are then fitted to the relation between requirement and norm and compared through information criteria: a linear model, a smooth nonlinear model such as a generalized additive model, and a segmented model with a freely estimated breakpoint. Support for the proposition requires that the segmented model be preferred and that the estimated breakpoint fall within a narrow interval across classes, allowing for the variation in class composition that the framework itself predicts. Hysteresis is tested by comparing matched requirement levels across the ascending and

Table 6. Testable propositions derived from the coupled-system model

Proposition	Core claim	Key constructs	Level	Suggested design and analysis
P1. Authority effect	The perceived credibility of generated output varies with linguistic fluency independently of factual accuracy, and the effect is stronger among students holding naive views of the nature of science	Perceived credibility; linguistic fluency; views of the nature of science	Individual	Experimental vignette study; fluency-by-accuracy manipulation; moderation analysis
P2. Materiality	Instructional sequences positioning model output as a hypothesis to be tested yield greater gains in understanding the nature of science than sequences relying on explanation alone, with hands-on time held constant	Instructional positioning; understanding of the nature of science	Individual, within class	Quasi-experimental comparison; pre-test and post-test measures
P3. Order parameter	Classroom-level verification norms predict individual verification behavior beyond individual attitudes	Verification norm; verification behavior; epistemic vigilance	Class and individual	Multilevel design; network psychometrics of classroom epistemic discourse
P4. Mode of engagement	The relation between model use and inquiry engagement depends on the mode of reliance rather than on its frequency	Dependent reliance; autonomous reliance; inquiry engagement	Individual, repeated measures	Repeated measures with mode and engagement measured at each wave; latent profile analysis as a secondary descriptive analysis
P5. Control parameter and phase transition	Graded increase in the teacher-set verification requirement produces discontinuous change in the classroom norm, and the norm is higher at a given requirement level on the descending than on the ascending path	Teacher-set requirement; classroom verification norm; hysteresis	Class, repeated measures	At least seven waves including a withdrawal phase and a constant-requirement comparison condition; comparison of linear, smooth nonlinear, and segmented models; breakpoint estimation

descending paths, since a difference between paths at the same level is what distinguishes hysteresis from the simple persistence a linear account also predicts. Classes assigned to a comparison condition in which the requirement is held constant throughout are needed for the same reason, so that persistence attributable to the passage of time can be separated from persistence attributable to the transition. Rising variance and temporal autocorrelation near the critical value can be examined as early-warning indicators, and across enough classes held at intermediate requirement levels the framework predicts a bimodal rather than a centered distribution. A longitudinal label alone does not specify this test; the number of waves, the withdrawal phase, and the comparison of functional forms are what render the threshold claim falsifiable. **Table 6** summarizes the five propositions together with their constructs and their proposed designs.

Instrumentation

Testing these propositions requires measures that do not yet exist as a set. A questionnaire based on the five categories of the FRA has recently been developed and used in a structural equation model of students' perceptions of generative-AI-influenced science (Cheung & Zhang, 2025). That instrument measures

perceptions rather than the classroom-level constructs the present framework requires. What remains to be developed is a matched set: an observation protocol for verification acts, a perceived descriptive norm scale at the classroom level, and a fidelity checklist for the teacher-set requirement, with evidence of measurement invariance for cross-cultural comparison. Developing these instruments is a prerequisite for the program above rather than a consequence of it.

Limitations

The design is conceptual; the proposed relations have not been tested but have been formulated so that they can be. The comparison depends on the architectures available at the time of writing and will require revision for systems that acquire the capacity for evidential interaction. The synergetic concept set has been used as a structured analogy, and its explanatory power is limited by the operationalization of order parameter and threshold within classroom data. The collapse of the social-institutional dimensions of the FRN into a single row, noted before, leaves the political economy of model development outside the analysis. It is not possible to make a generalization about the explanatory reach of the framework on conceptual grounds alone. Within these

limits, the framework is offered as an open starting point for an empirical program.

CONCLUSION

Text-based conversational systems present science education with a fluent source of answers that diverges from science in every category by which the mode of knowledge production in science is defined. This study is an attempt to reveal the structure of that divergence and to propose a constructive approach to teaching in the light of it. The family resemblance comparison identifies where the differences lie. The synergetic reframing accounts for them through a single contrast between an open system grounded in empirical evidence and a closed system that reorganizes a fixed body of knowledge, while the boundary cases keep that contrast at the level of the system rather than of the individual act. The coupled-system model translates the contrast into the language of design. In that model, classroom-level epistemic vigilance is the order to be established, the teacher-set verification requirement is the parameter that establishes it, and evidential accountability to the material world is the bond that keeps the arrangement open. The framework is presented as a conceptual foundation, and its value will be determined by the empirical program it enables.

Author contributions: ZOM: conceptualization, investigation, methodology, writing – original draft, writing – review & editing; DMM: investigation, visualization, writing – original draft, writing – review & editing; RIT: conceptualization, methodology, visualization; NTR: investigation, methodology, writing – original draft; OSOS: conceptualization, visualization, writing – review & editing; SUZ: investigation, writing – original draft, writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: This study is conceptual in nature; as it does not involve human participants or animal data, it does not require ethics committee approval.

AI statement: During the preparation of this study, generative AI tool (Claude Opus 5) for language editing and for checking the clarity of the text. After using it, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Abd-El-Khalick, F., Summers, R., Brunner, J. L., Belarmino, J., & Myers, J. (2024). Development of VAScoR: A rubric to qualify and score responses to the views of nature of science (VNOS) questionnaire. *Journal of Research in Science Teaching*, 61(7), 1641-1688. <https://doi.org/10.1002/tea.21916>
- Agustian, H. Y., Finne, L. T., Jorgensen, J. T., Pedersen, M. I., Christiansen, F. V., Gammelgaard, B., & Nielsen, J. A. (2022). Learning outcomes of university chemistry teaching in laboratories: A systematic review of empirical literature. *Review of Education*, 10(2), Article e3360. <https://doi.org/10.1002/rev3.3360>
- Alvarado, R. (2023). AI as an epistemic technology. *Science and Engineering Ethics*, 29, Article 32. <https://doi.org/10.1007/s11948-023-00451-3>
- Avraamidou, L. (2024). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*, 61(10), 2570-2574. <https://doi.org/10.1002/tea.21961>
- Barzilai, S., & Chinn, C. A. (2024). The AIR and Apt-AIR frameworks of epistemic performance and growth: Reflections on educational theory development. *Educational Psychology Review*, 36, Article 91. <https://doi.org/10.1007/s10648-024-09927-5>
- Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54(5), 1160-1173. <https://doi.org/10.1111/bjet.13337>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Binz, M., Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C. T., Allen, C., Schad, D., Wulff, D., West, J. D., Zhang, Q., Shiffrin, R. M., Gershman, S. J., Popov, V., Bender, E. M., Marelli, M., Botvinick, M. M., Akata, Z., & Schulz, E. (2025). How should the advancement of large language models affect the practice of science? *PNAS*, 122(5), Article e2401227121. <https://doi.org/10.1073/pnas.2401227121>
- Birhane, A., Kasirzadeh, A., Leslie, D., & Wachter, S. (2023). Science in the age of large language models. *Nature Reviews Physics*, 5(5), 277-280. <https://doi.org/10.1038/s42254-023-00581-4>
- Canty, R. B., Bennett, J. A., Brown, K. A., Buonassisi, T., Kalinin, S. V., Kitchin, J. R., Maruyama, B., Moore, R. G., Schrier, J., Seifrid, M., Sun, S., Vegge, T., & Abolhasani, M. (2025). Science acceleration and accessibility with self-driving labs. *Nature Communications*, 16, Article 3856. <https://doi.org/10.1038/s41467-025-59231-1>
- Cheung, K. K. C., & Erduran, S. (2023). A systematic review of research on the family resemblance approach to the nature of science in science education. *Science & Education*, 32(5), 1637-1673. <https://doi.org/10.1007/s11191-022-00379-3>

- Cheung, K. K. C., & Zhang, W. (2025). Undergraduate students' perceptions of the nature of generative artificial intelligence (GenAI)-influenced science: A structural equation modelling approach. *Journal of Science Education and Technology*, 34(6), 1495-1507. <https://doi.org/10.1007/s10956-025-10230-3>
- Cheung, K. K. C., Long, Y., Liu, Q., & Chan, H.-Y. (2025a). Unpacking epistemic insights of artificial intelligence (AI) in science education: A systematic review. *Science & Education*, 34(2), 747-777. <https://doi.org/10.1007/s11191-024-00511-5>
- Cheung, K. K. C., Zerouali, A., Koenen, J., & Erduran, S. (2025b). Do generative artificial intelligence (GenAI) and science education mix? A systematic review of the literature. *Studies in Science Education*. <https://doi.org/10.1080/03057267.2025.2578091>
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444-452. <https://doi.org/10.1007/s10956-023-10039-y>
- Du, Y., & Yuan, Y. (2026). Epistemic dependence in AI-mediated learning. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-026-03294-1>
- Erduran, S., & Dagher, Z. R. (2014). *Reconceptualizing the nature of science for science education: Scientific knowledge, practices, and other family categories*. Springer. <https://doi.org/10.1007/978-94-017-9057-4>
- Erduran, S., Dagher, Z. R., & McDonald, C. V. (2019). Contributions of the family resemblance approach to the nature of science in science education. *Science & Education*, 28(3-5), 311-328. <https://doi.org/10.1007/s11191-019-00052-2>
- Favela, L. H., & Amon, M. J. (2023). Reframing cognitive science as a complexity science. *Cognitive Science*, 47(4), Article e13280. <https://doi.org/10.1111/cogs.13280>
- Ferrario, A., Facchini, A., & Termine, A. (2024). Experts or authorities? The strange case of the presumed epistemic superiority of artificial intelligence systems. *Minds and Machines*, 34, Article 30. <https://doi.org/10.1007/s11023-024-09681-1>
- Gao, Y., Zhai, X., Li, M., Lee, G., & Liu, X. (2025). A multimodal interactive framework for science assessment in the era of generative artificial intelligence. *Journal of Research in Science Teaching*, 62(9), 2014-2028. <https://doi.org/10.1002/tea.70009>
- Garofalo, S. G., & Farenga, S. J. (2025). Science teacher perceptions of the state of knowledge and education at the dawn of generative artificial intelligence's rise in popularity. *Science & Education*, 34(2), 893-912. <https://doi.org/10.1007/s11191-024-00534-y>
- Gericke, N., Hogstrom, P., & Wallin, J. (2023). A systematic review of research on laboratory work in secondary school. *Studies in Science Education*, 59(2), 245-285. <https://doi.org/10.1080/03057267.2022.2090125>
- Gyamfi, M., & Ayaaba, D. A. (2026). Understanding the nature of science in biology education: A meta-analysis of students' and teachers' views in senior high schools. *Eurasian Journal of Science and Environmental Education*, 6(1), 1-5. <https://doi.org/10.30935/ejsee/17753>
- Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511814563>
- Haken, H. (2004). *Synergetics: Introduction and advanced topics*. Springer. <https://doi.org/10.1007/978-3-662-10184-1>
- Han, M., & Gutierrez, S. B. (2026). "How can we check the evidence presented by GenAI's response?": Students' epistemic vigilance of their co-constructed scientific explanations with GenAI. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2026.2717405>
- He, J., & Yang, C. (2025). Testimony by LLMs. *AI & SOCIETY*, 40(8), 6201-6213. <https://doi.org/10.1007/s00146-025-02366-y>
- Heersmink, R., de Rooij, B., Clavel Vazquez, M. J., & Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. *Ethics and Information Technology*, 26, Article 41. <https://doi.org/10.1007/s10676-024-09777-3>
- Hetherington, L., Hardman, M., Noakes, J., & Wegerif, R. (2018). Making the case for a material-dialogic approach to science education. *Studies in Science Education*, 54(2), 141-176. <https://doi.org/10.1080/03057267.2019.1598036>
- Hila, A. (2026). The epistemological consequences of large language models: Rethinking collective intelligence and institutional knowledge. *AI & SOCIETY*, 41, 79-97. <https://doi.org/10.1007/s00146-025-02426-3>
- Huwer, J., Thyssen, C., Becker-Genschow, S., von Kotzebue, L., Finger, A., Kremser, E., Berber, S., Brückner, M., Maurer, N., Buckermann, T., Meier, M., & Thoms, L. (2025). Competencies for teaching with and about artificial intelligence in the natural sciences: DiKoLAN AI. *Computers and Education Open*, 9, Article 100303. <https://doi.org/10.1016/j.caeo.2025.100303>
- Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education*, 20(7-8), 591-607. <https://doi.org/10.1007/s11191-010-9293-4>

- Ishmuradova, I., Zhdanov, S. P., Boldyreva, N. V., Nekrasova, I. E., Vasyukov, V. F., & Kondakchian, N. A. (2025). Pre-service science teachers' perceptions of using generative artificial intelligence in science education. *Contemporary Educational Technology*, 17(3), Article ep579. <https://doi.org/10.30935/cedtech/16207>
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1-2), 18-26. <https://doi.org/10.1007/s13162-020-00161-0>
- Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., Mumthas, S., & Joseph, S. (2025). Epistemic authority and generative AI in learning spaces: Rethinking knowledge in the algorithmic age. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1647687>
- Kaya, E., & Erduran, S. (2016). From FRA to RFN, or how the family resemblance approach can be transformed for science curriculum analysis on the nature of science. *Science & Education*, 25(9-10), 1115-1133. <https://doi.org/10.1007/s11191-016-9861-3>
- Koopmans, M. (2020). Education is a complex dynamical system: Challenges for research. *The Journal of Experimental Education*, 88(3), 358-374. <https://doi.org/10.1080/00220973.2019.1566199>
- Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., dos Passos Gomes, G., Hase, F., Jinich, A., Nigam, A., Yao, Z., & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12), 761-769. <https://doi.org/10.1038/s42254-022-00518-3>
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497-521. <https://doi.org/10.1002/tea.10034>
- Lee, G., Latif, E., Nyaaba, M., Gao, Y., Guo, S., Shi, L., Bewersdorff, A., Cooper, G., Ha, M., Feldman-Maggor, Y., Tai, R. H., Tang, K.-S., Chng, E., Nerdel, C., & Tang, K.-S. (2026). Transforming science teacher education in the era of generative artificial intelligence: Global perspectives. *International Journal of Science Education*, 48(15), 2523-2552. <https://doi.org/10.1080/09500693.2026.2614365>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktaschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Liu, X., Chen, T., Da, L., Chen, C., Lin, Z., & Wei, H. (2025). Uncertainty quantification and confidence calibration in large language models: A survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 6107-6117). ACM. <https://doi.org/10.1145/3711896.3736569>
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>
- Martin-Moncunill, D., & Alonso Martinez, D. (2025). Students' trust in AI and their verification strategies: A case study at Camilo Jose Cela University. *Education Sciences*, 15(10), Article 1307. <https://doi.org/10.3390/educsci15101307>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.
- Morin, E. (2008). *On complexity*. Hampton Press.
- Muminova, Z. O. (2025). Synergetic thinking and artificial intelligence: A philosophical analysis. *International Journal of History and Political Sciences*, 5(12), 62-65. <https://doi.org/10.37547/ijhps/Volume05Issue12-14>
- Olaniyan, Y. D., Martins, M. O., & Al Maqrashi, R. H. (2026). Generative artificial intelligence and epistemic (in)justice: Perspectives from higher education students in the Global North and South. *Frontiers in Human Dynamics*, 8. <https://doi.org/10.3389/fhumd.2026.1790324>
- Osborne, J., & Pimentel, D. (2023). Science education in an age of misinformation. *Science Education*, 107(3), 553-571. <https://doi.org/10.1002/sce.21790>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730-27744. <https://doi.org/10.52202/068431-2011>
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226668253.001.0001>
- Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson. <https://doi.org/10.1063/1.3060577>
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. Bantam.

- Rivera-Novoa, A., & Duarte Arias, D. A. (2025). Generative artificial intelligence and extended cognition in science learning contexts. *Science & Education*, 35(3), 755-776. <https://doi.org/10.1007/s11191-025-00660-1>
- Sengul, O. (2023). Pre-service mathematics teachers' views of nature of science in the context of COVID-19. *European Journal of Science and Mathematics Education*, 11(3), 499-514. <https://doi.org/10.30935/scimath/12982>
- Sevryugina, Y. V., & Vargas, D. (2026). Teaching research integrity through verification of AI-generated references: An activity for upper-level chemistry courses. *Journal of Chemical Education*, 103(5), 2610-2620. <https://doi.org/10.1021/acs.jchemed.5c01620>
- Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25(4), 359-393. <https://doi.org/10.1111/j.1468-0017.2010.01394.x>
- Steyvers, M., Tejada, H., Kumar, A., Belem, C., Karny, S., Hu, X., Mayer, L. W., & Smyth, P. (2025). What large language models know and what people think they know. *Nature Machine Intelligence*, 7(2), 221-231. <https://doi.org/10.1038/s42256-024-00976-7>
- Suazo Galdames, I. C., Santiesteban Velazquez, M., Saracostti, M., & Chaple Gil, A. M. (2026). Between cognitive offloading and critical autonomy: A systematic review of the epistemic implications of generative AI in higher education. *Frontiers in Education*, 11. <https://doi.org/10.3389/feduc.2026.1878665>
- Tang, K.-S. (2022). Material inquiry and transformation as prerequisite processes of scientific argumentation: Toward a social-material theory of argumentation. *Journal of Research in Science Teaching*, 59(6), 969-1009. <https://doi.org/10.1002/tea.21749>
- Tang, K.-S. (2024). Informing research on generative artificial intelligence from a language and literacy perspective: A meta-synthesis of studies in science education. *Science Education*, 108(5), 1329-1355. <https://doi.org/10.1002/sce.21875>
- Tang, K.-S. (2026). Reconceptualizing generative AI as dialogic partner: Pedagogical and methodological implications for science education. *International Journal of Science and Mathematics Education*, 24, Article 89. <https://doi.org/10.1007/s10763-026-10710-x>
- Tang, K.-S., & Cooper, G. (2025). The role of materiality in an era of generative artificial intelligence. *Science & Education*, 34(2), 731-746. <https://doi.org/10.1007/s11191-024-00508-0>
- Tom, G., Schmid, S. P., Baird, S. G., Cao, Y., Darvish, K., Hao, H., Lo, S., Pablo-Garcia, S., Rajaonson, E. M., Skreta, M., Yoshikawa, N., Corapi, S., Akkaoc, G. D., Strieth-Kalthoff, F., Seifrid, M., & Aspuru-Guzik, A. (2024). Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 124(16), 9633-9732. <https://doi.org/10.1021/acs.chemrev.4c00055>
- Urban, M., Brom, C., Lukavsky, J., Dechterenko, F., Hein, V., Svacha, F., Kmonickova, P., & Urban, K. (2025). "ChatGPT can make mistakes. Check important info." Epistemic beliefs and metacognitive accuracy in students' integration of ChatGPT content into academic writing. *British Journal of Educational Technology*, 56(5), 1897-1918. <https://doi.org/10.1111/bjet.13591>
- van Dijk, E. M. (2011). Portraying real science in science communication. *Science Education*, 95(6), 1086-1100. <https://doi.org/10.1002/sce.20458>
- Wang, S., & Zhang, H. (2026). Pedagogical partnerships with generative AI in higher education: How dual cognitive pathways paradoxically enable transformative learning. *International Journal of Educational Technology in Higher Education*, 23, Article 11. <https://doi.org/10.1186/s41239-026-00585-x>
- Wickman, P.-O. (2004). The practical epistemologies of the classroom: A study of laboratory work. *Science Education*, 88(3), 325-344. <https://doi.org/10.1002/sce.10129>
- Woodruff, E., & Hewitt, J. (2026). Epistemic agency in the age of large language models: Design principles for knowledge-building AI. *AI*, 7(3), Article 99. <https://doi.org/10.3390/ai7030099>
- Xu, C., & Fan, D. (2024). Creative-becoming holism: Reflections on and development of creative holism in the case of science education. *Systems Research and Behavioral Science*, 42(1), 98-110. <https://doi.org/10.1002/sres.3106>
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458-477. <https://doi.org/10.5951/jresmetheduc.27.4.0458>
- Yampolskiy, R. V. (2020). Unexplainability and incomprehensibility of AI. *Journal of Artificial Intelligence and Consciousness*, 7(2), 277-291. <https://doi.org/10.1142/S2705078520500150>
- Yan, L., Greiff, S., Teuber, Z., & Gasevic, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839-1850. <https://doi.org/10.1038/s41562-024-02004-5>

- Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology*, 34(2), 265-288. <https://doi.org/10.1007/s13347-019-00382-7>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhu, Q., Li, X., Dong, Y., Chang, P., & Fan, M. (2026). Not all cognitive offloading is equal: Distinguishing dependent and autonomous offloading to generative AI. *Frontiers in Psychology*, 17. <https://doi.org/10.3389/fpsyg.2026.1878629>
- Zhuo, R., & Bai, Z. (2026). Mitigating cognitive offloading in GenAI-supported collaboration: A metacognitive scaffolding approach to enhancing students' critical thinking. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2026.2697043>

<https://www.ejmste.com>