International Journal of Multidisciplinary Research and Growth Evaluation.

Integrating Generative AI into STEAM Geography Education: A Framework for Evidence-Based Reasoning

Viktoria Dimitrova
Faculty of Geology and Geography, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

* Corresponding Author: **Viktoria Dimitrova**

Article Info

ISSN (Online): 2582-7138
Volume: 07
Issue: 05
Received: 22-07-2026
Accepted: 26-08-2026
Published: 24-09-2026
Page No: 549-554

Abstract
Background: Generative artificial intelligence can supply geographical explanations before students have examined the evidence needed to evaluate them. STEAM activities offer a setting in which such explanations can be tested against observations, representations and disciplinary knowledge.
Objectives: To develop a framework for integrating generative AI into secondary geography while retaining student responsibility for scientific reasoning.
Methods: A targeted narrative synthesis of literature on spatial thinking, integrated STEM and STEAM, cognitive engagement and AI in education informed a conceptual design. A previously published geography–physics lesson provided the instructional context; no new student data were collected.
Results: The resulting framework connects five stages: initial explanation, inspection of AI output, evidence checking, model revision and independent justification. An illustrative 80-minute lesson on atmospheric circulation specifies student activities, teacher responsibilities, verification procedures and assessment criteria. Its central design requirement is that substantive claims retained from AI output must be supported by identifiable evidence or explicitly qualified.
Conclusion: The framework provides a testable proposal for using AI-generated explanations as material for geographical inquiry. Its educational effects remain to be established through classroom implementation and assessment of independent student reasoning.

DOI: https://doi.org/10.54660/.IJMRGE.2026.7.5.549-554

Keywords: Atmospheric circulation, Evidence-based reasoning, Generative artificial intelligence, Geography education, STEAM education

1. Introduction

Generative artificial intelligence is becoming increasingly embedded in the digital environments used by secondary-school students, making its complete exclusion from formal learning progressively less realistic. The pedagogical question is therefore no longer simply whether AI should enter the classroom, but under what conditions its use can remain limited, transparent and subordinate to students' own reasoning. This creates a need for instructional models that position AI as a bounded support and an object of critical evaluation rather than as a substitute for disciplinary thinking.

A geography lesson on atmospheric circulation asks students to connect processes that operate at different scales. They must relate a local observation of convection to a global representation, distinguish prevailing winds from weather on a particular day, and explain why a spatial pattern is plausible. Spatial literacy includes evaluating geographical information and using it to defend a line of reasoning [1]. A labelled diagram or fluent paragraph therefore provides only part of the evidence that a student understands the topic.

Generative artificial intelligence (AI) introduces an additional source of explanations into this setting. Students can obtain a coherent answer without having decided which observations would support it. UNESCO's guidance identifies the need to examine the pedagogical suitability of generative AI and preserve human agency [2]. For geography teaching, this raises a practical question: how can work with generated explanations require students to exercise disciplinary judgement?

A published Bulgarian study described an interdisciplinary geography–physics STEAM lesson on general atmospheric circulation, combining experimentation, data interpretation and model construction [3]. That instructional combination provides the starting point for the present design. The earlier classroom results concern the intervention reported in that publication. They are neither reanalysed here nor treated as evidence about generative AI.

This article develops a conceptual framework in which AI output becomes an object of investigation. The guiding question is: how can generative AI be incorporated into a STEAM geography lesson while preserving students' responsibility for selecting evidence, explaining relationships and justifying revisions? The contribution consists of a sequence of activities; a verification procedure and assessment criteria illustrated through atmospheric circulation. These are proposed design choices requiring future evaluation, rather than findings from a new classroom experiment.

2 Theoretical backgrounds

2.1. Integration Through A Shared Geographical Problem

Kelley and Knowles connect integrated STEM education with situated learning, scientific inquiry, engineering design, technological literacy and mathematical thinking [4]. Herro and colleagues address STEAM collaboration and develop an instrument for assessing students' contributions to joint problem solving [5]. Together, these perspectives suggest that an interdisciplinary lesson needs a shared problem and observable intellectual contributions. Simply using several tools does not establish a connection between the subject areas.

The present framework places the explanation of a geographical pattern at the centre of the activity. Scientific concepts constrain the explanation; numerical comparisons help students examine it; and visual design makes the proposed relationships inspectable. Engineering is represented by designing, testing and revising a physical or digital explanatory model against specified requirements. The arts component concerns decisions about visual communication and audience understanding. These functions must remain explicit when AI is added, otherwise the task can become an exercise in obtaining and decorating generated text.

The lesson is treated as a short project-oriented sequence rather than a complete extended project-based learning programme. Kokotsaki and colleagues identify student autonomy, investigation, collaboration and reflection as features of project-based learning, while also emphasising scaffolding and aligned assessment [6]. In the proposed sequence, students choose how to represent their explanation, but the teacher supplies a manageable evidence set and a clear standard for accepting claims. This balance is particularly relevant when the class has limited time or uneven experience of reading scientific sources.

2.2. Visible Reasoning and Individual Understanding

The ICAP framework distinguishes passive, active, constructive and interactive modes of engagement [7]. It offers a useful lens for considering what learners actually do with AI output. Highlighting a paragraph, producing an additional explanation and jointly resolving a disagreement are different activities. The present application of ICAP is a design interpretation; the framework does not establish that interaction with a chatbot is equivalent to productive dialogue between learners.

Scientific explanation also requires a connection between claims and evidence. McNeill and colleagues investigated support for students' construction of explanations using claims, evidence and reasoning [8]. Here, that distinction informs a short verification record: the claim being examined, the evidence consulted and the reason for retaining or changing it. A source address alone is insufficient. Students need to identify what the source shows and how it bears on the claim.

Individual understanding must remain visible within group work. A model can be scientifically sound because one student corrected it, while others still hold incompatible explanations. The proposed lesson therefore pairs a shared product with an individual initial explanation and a final task completed without AI. These records are intended to reveal changes in reasoning that the finished model might conceal. They also give the teacher a basis for responding to a particular misconception rather than assigning the same judgement to every member of a group.

2.3. Ai Literacy and The Limits of Assistance

UNESCO's student framework includes a human-centred mindset, AI ethics, techniques and applications, and system design [9]. Its teacher framework addresses AI foundations, pedagogy and professional learning alongside ethical and human-centred considerations [10]. The present proposal concentrates on a narrow part of this wider competence: evaluating generated geographical content and making accountable decisions about its use. It does not constitute a complete AI literacy curriculum.

The distinction between assisted performance and learning is supported by evidence outside geography. In a randomised study of high-school mathematics, Bastani and colleagues found that a general AI interface improved performance during supported practice but reduced subsequent unassisted performance relative to the control group. A tutor with learning-oriented safeguards largely mitigated the negative effect [11]. These findings should not be transferred directly to atmospheric circulation. They justify treating tool design and independent assessment as research questions rather than assuming that a better group product demonstrates stronger learning.

Accordingly, AI assistance is deliberately limited in this framework. It may provide a short explanation for critique, ask a question or suggest an alternative representation. The students still decide what counts as evidence and construct the final justification. The teacher checks scientific suitability and determines when further interaction would support the task and when it would merely supply an answer.

3. Materials and Methods

This is a conceptual and methodological article based on targeted narrative synthesis and instructional design. It reports no new participants, intervention outcomes or statistical analyses. Sources were located through targeted web searches and checks of publisher, university repository and institutional pages. Search terms combined STEAM, integrated STEM, geography education, spatial thinking, scientific explanation, cognitive engagement, generative AI and AI competency frameworks. The author's previously published Bulgarian study was examined as an existing instructional case.

Selection was purposeful rather than exhaustive. Publications were retained when they supplied a clear rationale for a design decision, an assessment construct, relevant empirical evidence or an authoritative explanation of the geographical content. Foundational educational literature was considered alongside recent AI research and UNESCO guidance. The article makes no claim to a systematic review, a comprehensive database search or a formal quality appraisal of all relevant research. Its literature base supports the proposed design but cannot establish the prevalence or overall effectiveness of AI-supported geography teaching.

The synthesis proceeded through three design questions: what intellectual work should remain with the learner; what evidence can make that work visible; and how can the activity fit the constraints of a secondary-school lesson? Spatial reasoning, interdisciplinary integration, explanation, collaboration and human oversight were used as organising concepts. Each proposed stage was connected to an observable student action and a record that could be assessed. This procedure produced the framework in Table 1 and the illustrative sequence in Section 5.

The existing atmospheric-circulation lesson [3] contributed the subject context and the combination of experimentation, representation and group work. The new contribution is the structured examination of AI output, a record of evidence-based revisions and an independent explanation task. Numerical outcomes, diagrams and assessment results from the earlier publication are not reproduced. The illustrative prompts and assessment descriptors below were developed for this proposal and have not been validated.

Scientific content requires its own verification. The Met Office account of global circulation provides an accessible reference for the three-cell representation and the role of Earth's rotation [12]. ERA5 monthly data offer a possible resource for preparing geographical comparisons; they are reanalysis estimates that combine observations and modelling, not a set of direct classroom measurements [13]. These resources are identified for a future lesson. No ERA5 data were downloaded or analysed for the present article, and no generated numerical values are presented as observations. As this article reports no new research involving human participants, personal data or animals, ethical approval and informed consent were not applicable.

4. Results: Proposed Instructional Framework

4.1. A Record of Decisions

The framework begins before AI is introduced. Each student writes an initial explanation of the geographical problem and marks one point of uncertainty. This creates a record of prior reasoning and prevents the generated answer from being the only account available for discussion. Initial answers are formative and should not be penalised for containing ideas that the lesson is designed to examine.

The teacher then provides a short, previously checked AI response or supervises a bounded interaction. Students separate the response into claims that can be examined individually. They classify each as supported, requiring qualification, contradicted or unresolved. These categories describe the current judgement about a claim; they are not labels for the reliability of an entire tool. A correct claim should be retained when the evidence supports it, just as an unsupported claim should be challenged.

For every substantive claim retained in the product, the group records the relevant source or observation and a brief reason for accepting it. Where the evidence is insufficient, the group narrows the statement or records uncertainty. This requirement applies equally to student-generated and AI-generated claims. It is intended to make evaluation a normal part of geographical inquiry rather than a search for mistakes in a particular technology.

Revision must change the explanation when necessary. Students identify an original claim, the evidence that prompted reconsideration and the corresponding change in their model or written account. A revision that only improves grammar does not demonstrate conceptual progress. The final stage asks students to defend a decision individually, without access to the chatbot or the group's completed paragraph.

Table 1: Integrating generative AI into a STEAM geography lesson through evidence and assessment

Stage	Student activity	Role of generative AI	Verification	Assessable evidence
1 Initial explanation	Explain a spatial pattern and sketch the expected relationships.	Not used at this stage.	Compare initial ideas with the lesson question and known concepts.	Individual explanation and a stated uncertainty.
2 Inspect output	Separate a short, generated explanation into checkable claims.	Provides a bounded text selected by the teacher.	Identify terms, causal statements and assumptions requiring a check.	Annotated text with reasons for selecting claims.
3 Check evidence	Compare claims with sources, data and a physical demonstration.	May ask a clarifying question; does not certify a claim.	Record source, relevant evidence, scale and any limits.	Claim-by-claim judgement supported by a brief justification.
4 Revise the model	Build or modify a representation and document a substantive change.	May suggest a representation or question the clarity of the explanation.	Test the product against scientific and communication criteria.	Revised model and an explicit link between evidence and revision.
5 Justify independently	Explain one correction and address a new but related question.	Not used during individual assessment.	Teacher compares the explanation with subject criteria.	Individual response distinguishing evidence from assertion.

Source: Author's proposed synthesis. The sequence is an instructional design, not an empirically validated intervention.

4.2. Teacher Oversight and Access

Teachers prepare the evidence set and review the generated text before the lesson. Each extract is retained with the tool name, displayed model identifier where available, generation date and prompt. Teacher edits are recorded separately so that an intentionally introduced error is not misrepresented as a naturally occurring model response. This record supports later interpretation if the lesson is evaluated and helps distinguish changes in teaching design from changes in model output.

A printed response can serve the same instructional purpose as a live chatbot when the objective is to evaluate claims. This option gives all groups access to the same text and reduces variation caused by connectivity or different accounts. It also permits the lesson to proceed without requiring personal student accounts. A live interaction is an optional extension, subject to school procedures and applicable service conditions; it is not a prerequisite for participation.

The teacher's role is especially important when a plausible simplification becomes a misconception. For example, a classroom convection demonstration can illustrate buoyancy-driven movement, but it does not reproduce Earth's rotation or the full dynamics of global circulation. Students should state which feature the demonstration represents and which features require another source of explanation. The teacher uses such distinctions to decide whether a disagreement can be resolved by observation, by consulting a scientific reference or by acknowledging the limits of the classroom model.

5. Illustrative Lesson on Atmospheric Circulation

5.1. Context and Preparation

The lesson sequence presented below adapts the instructional context of the previously published geography–physics STEAM lesson [3]. The present design substantially changes its pedagogical purpose by introducing structured inspection of AI-generated claims, evidence-based verification, documented revision and independent justification. The earlier intervention and its empirical outcomes are not reproduced.

The proposed lesson is designed for Grade 9 students who have already encountered temperature, pressure, latitude and the water cycle. It occupies an 80-minute block and is intended for application and consolidation. Introducing all the underlying concepts for the first time would require additional lessons. Geography and physics teachers may teach it jointly; one teacher could also implement it after checking the scientific content and preparing the demonstration.

The design problem is to produce a small explanatory display for younger pupils: why are rising air and frequent rainfall associated with the equatorial belt, while descending air is associated with many subtropical dry regions? Groups must explain the relationship and identify one limitation of their model. The problem gives the final product an audience and a purpose. It also makes choices about symbols, arrows, captions and uncertainty part of the reasoning task.

Before class, the teacher prepares a concise scientific account, a labelled reference diagram and a small data extract with its provenance. The data should represent an appropriate averaging period rather than a single weather snapshot. For instance, a prepared ERA5 extract could compare monthly climatological conditions across selected latitude bands [13]. Its caption must name the variables, units, geographical coverage, averaging period and processing method. Students receive the prepared material; acquiring and processing global datasets is outside the 80-minute allocation.

The teacher also prepares one short AI response using a prompt such as: Explain the relationship between rising air, pressure and rainfall near the equator for Grade 9 students. Distinguish a long-term circulation pattern from weather on a particular day. This is a proposed prompt, not a transcript of an interaction used in a completed study. If the response contains no useful ambiguity, students can still evaluate its scope and evidential basis. There is no requirement to manufacture an AI error.

5.2. An 80-Minute Sequence

During the first 10 minutes, students answer the display question individually and sketch the expected direction of vertical air movement. They then identify the part of their explanation in which they are least confident. The teacher collects or photographs these responses before distributing the generated text. This ordering protects the initial account from being reconstructed after students have seen a model answer.

The next 10 minutes are used to inspect the AI response in groups of three or four. Students mark two or three substantive claims and identify what evidence could test each. Roles include coordinating the discussion, checking sources, examining the representation and recording decisions, with roles combined in smaller groups. Each student must explain at least one judgement; assigning one permanent source checker would risk concentrating the reasoning in a single person.

A 20-minute evidence phase combines a teacher-led convection demonstration with examination of the reference materials. The demonstration uses a suitable transparent container and a controlled temperature contrast under normal school laboratory procedures. Students describe the observed motion before discussing its analogy with atmospheric movement. The scientific reference is used to check the role of rotation and the three-cell representation [12]. A short numerical comparison can involve differences between supplied mean values, with units and averaging periods explicitly recorded. The purpose is to relate a calculation to an explanation; no formula for wind speed is inferred from an isolated pressure difference.

During the following 25 minutes, groups construct and test their explanatory display. Reusable components or a prepared digital layout keep construction feasible within the block. The product must distinguish surface and upper-level movement, use a consistent legend, connect one numerical comparison to a geographical statement and include a limitation. Groups exchange displays and test whether another group can interpret the arrows and identify the evidence for the main claim. They then make one documented revision.

The remaining 15 minutes combine a brief comparison of selected decisions with an individual exit response. Instead of requiring every group to give a lengthy presentation, the teacher selects contrasting revisions for discussion. Each student then explains one claim they accepted or changed and why. A transfer question might ask whether a rainy day in a subtropical location disproves the proposed circulation explanation. Students must distinguish a particular event from an averaged pattern, using the reasoning developed during the lesson.

5.3. Assessment and Feedback

Assessment gives priority to scientific reasoning, with presentation quality judged in relation to meaning. Four proposed dimensions are used: scientific accuracy, relevance of evidence, explanation of causal relationships and justification of revision. For each dimension, a three-level descriptor can distinguish an absent or contradictory response, a partly supported response and a coherent response with an explicit justification. These descriptors are provisional and require piloting before they are used as a research instrument.

For evidence use, the lowest level would include a claim accompanied only by an AI assertion or an irrelevant source. An intermediate response would identify a relevant source but leave its connection to the claim unclear. The highest level would specify the relevant observation or information and explain why it supports an appropriately bounded claim. The same principle applies to revision: changing a label is weaker evidence of learning than explaining how a corrected relationship changes the model.

Group collaboration is documented separately from conceptual understanding. Co-Measure demonstrates the importance of assessing individual contributions within STEAM collaboration [5], but the present proposal does not claim to reproduce or validate that instrument. A short teacher record can note whether students request evidence, explain a disagreement and respond to a peer's reasoning. Such observations supplement the individual written task; they do not replace it.

Feedback identifies the next reasoning step. If a student cites a source without explaining its relevance, the teacher asks which part of the source bears on the claim. If the student generalises from the demonstration to the entire atmosphere, feedback asks which processes were absent from the demonstration. The assessment thus remains connected to the decisions made during the task, while preserving a distinction between an attractive product and an independently defensible explanation.

6. Discussion and Limitations

The proposed contribution is the explicit connection between AI use, evidence checking and assessment within a geography lesson. The framework specifies when generated content enters the activity, what students must do with it and which records make their decisions visible. Its instructional value would depend on whether those requirements produce substantive reasoning. A completed verification sheet, by itself, could become another routine that students perform without understanding.

Several tensions therefore deserve attention. Giving students a common AI response improves comparability but reduces the freedom available in an open conversation. Allowing live interaction may support questions that matter to a particular group, yet also changes the material that different students encounter. Similarly, a tightly selected source set makes evaluation manageable for novices but offers limited practice in finding and comparing unfamiliar sources. These choices should be matched to the learning objective and documented in any later evaluation.

The design also depends on subject knowledge. Students cannot evaluate every geographical claim through general scepticism alone. A novice may reject a correct explanation because it conflicts with an existing misconception. Teacher modelling, accessible references and an opportunity to leave a claim unresolved are therefore included in the proposal. The initial and final individual responses should help identify whether confidence has become better aligned with the quality of reasoning, rather than merely increased.

The short duration constrains what can be concluded. An 80-minute lesson can reveal how students approach a bounded task, but it cannot establish sustained AI literacy or transfer across the geography curriculum. The framework is presently limited to a physical-geography example. Adaptations to population, economic geography or environmental decision-making would require attention to different data, competing interpretations and value judgements. The earlier Bulgarian lesson supplies a relevant teaching context, not evidence that these adaptations will work.

A future pilot should first test feasibility: timing, source readability, the interpretability of the table and the consistency with which teachers apply the assessment descriptors. A subsequent comparative study could use the same STEAM task, teaching time and evidence materials in both conditions, with the structured AI activity added in one. Baseline knowledge, teacher differences and clustering by class would need to be addressed. If assignment is not random, the limits on causal interpretation should be explicit.

The main outcomes should include independent conceptual explanation and evidence use, assessed immediately and after a delay. Blind scoring by more than one assessor, with a check of scoring agreement, would strengthen interpretation. Group products, revision records and observations could explain how particular outcomes arose. Self-reported enjoyment would provide complementary information but could not establish that students learned more. A comparison with an equally plausible teacher-written explanation would further test whether any benefit derives from AI specifically or from the act of evaluating an external explanation.

Such a programme could also reveal situations in which AI adds little. If preparing and supervising the generated text increases workload without improving reasoning, a conventional source-comparison activity may serve the same purpose. The framework is intended to make that question testable. Its justification rests on the quality of the geographical inquiry it enables, not on a requirement to include AI in every lesson.

7. Conclusion

Generative AI can be incorporated into STEAM geography through a sequence that makes students accountable for the explanations they retain. The proposed framework combines an initial independent account, inspection of generated claims, evidence checking, revision of a model and individual justification. Its atmospheric-circulation example shows how those stages can be organised within an 80-minute consolidation lesson.

The article offers an instructional proposal and an assessment focus, not evidence of improved attainment. Classroom evaluation should determine whether the sequence strengthens independent reasoning, for whom it is feasible and whether AI adds value beyond other forms of source evaluation. Until those questions are examined, the framework is best treated as a structured basis for design, adaptation and empirical investigation.

Acknowledgements

The published lesson that informed the instructional context is acknowledged through its citation [3].

Declaration of Competing Interests

[AUTHOR TO CONFIRM: The author declares no competing interests.]

Use of Artificial Intelligence

ChatGPT (OpenAI) was used to assist with literature discovery, conceptual organisation, drafting in English and language editing. AI assistance was also used to formulate the proposed instructional activities and organise the table. The author retains responsibility for the scientific content and the final manuscript. No AI-generated student data or classroom outcomes are reported.

References

1. National Research Council. Learning to think spatially. Washington (DC): National Academies Press; 2006. doi:10.17226/11019.
2. Miao F, Holmes W. Guidance for generative AI in education and research. Paris: UNESCO; 2023. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000386693
3. Dimitrova V. Implementing the STEAM approach in an interdisciplinary geography and physics lesson: pedagogical effects in learning the general atmospheric circulation of the Earth. Probl Geogr. 2026;(1-2):143-159. Bulgarian. doi:10.7546/PG.1-2.2026.07.
4. Kelley TR, Knowles JG. A conceptual framework for integrated STEM education. Int J STEM Educ. 2016;3:11. doi:10.1186/s40594-016-0046-z.
5. Herro D, Quigley C, Andrews J, Delacruz G. Co-Measure: developing an assessment for student collaboration in STEAM activities. Int J STEM Educ. 2017;4:26. doi:10.1186/s40594-017-0094-z.
6. Kokotsaki D, Menzies V, Wiggins A. Project-based learning: a review of the literature. Improving Schools. 2016;19(3):267-277. doi:10.1177/1365480216659733.
7. Chi MTH, Wylie R. The ICAP framework: linking cognitive engagement to active learning outcomes. Educ Psychol. 2014;49(4):219-243. doi:10.1080/00461520.2014.965823.
8. McNeill KL, Lizotte DJ, Krajcik J, Marx RW. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. J Learn Sci. 2006;15(2):153-191. doi:10.1207/s15327809jls1502_1.
9. Miao F, Shiohira K, Lao N. AI competency framework for students. Paris: UNESCO; 2024. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000391105
10. Miao F, Cukurova M. AI competency framework for teachers. Paris: UNESCO; 2024. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000391104
11. Bastani H, Bastani O, Sungu A, Ge H, Kabakcı Ö, Mariman R. Generative AI without guardrails can harm learning: evidence from high school mathematics. Proc Natl Acad Sci U S A. 2025;122(26):e2422633122. doi:10.1073/pnas.2422633122.
12. Met Office. Global circulation patterns [Internet]. Exeter: Met Office; [cited 2026 Sep 2[2]]. Available from: https://weather.metoffice.gov.uk/learn-about/weather/atmosphere/global-circulation-patterns
13. Copernicus Climate Change Service. ERA5 monthly averaged data on single levels from 1940 to present [dataset]. Climate Data Store; 2019 [cited 2026 Sep 2[2]]. doi:10.24381/cds.f17050d7.

How to Cite This Article

Dimitrova V. Integrating generative AI into STEAM geography education: a framework for evidence-based reasoning. Int J Multidiscip Res Growth Eval. 2026;7(5):549-554. doi:10.54660/IJMRGE.2026.7.5.549-554.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.