THINKING BEFORE PROMPTING: LINKING SPATIAL REASONING, CRITICAL THINKING AND GENERATIVE AI THROUGH THE SPARC FRAMEWORK

Ergi BUFASI [1,2*]

[1] *Department of Physics, Rīga Stradiņš University, Riga, Latvia*
[2] *Institute of Physics and Materials Science, Riga Technical University, Riga, Latvia*
[*] *Corresponding author e-mail: ergi.bufasi@rsu.lv*

Abstract

Generative artificial intelligence (GenAI) produced explanations, diagrams and worked solutions in seconds, which raised a pedagogical dilemma: better task completion did not show that learners had practised the reasoning needed for independent performance. This conceptual paper examined the intersection of spatial reasoning, scientific reasoning and critical thinking in AI-supported higher education. A focused, illustrative synthesis of research on spatial cognition, cognitive offloading, human-AI interaction and educational scaffolding identified a plausible mechanism of concern: when AI supplied both a representation and its interpretation, students could bypass the construction, transformation and checking of mental models. The paper distinguished supportive from substitutive offloading and proposed SPARC (Sketch, Predict, Audit, Reconstruct, Calibrate), a five-stage instructional sequence that preserved learner-generated spatial and causal reasoning while using AI as a fallible comparison partner. Two teaching vignettes from introductory physics, the vertical motion of a tossed ball and work and power in stair climbing, showed how the sequence could be implemented with brief written or visual artefacts and assessed with an illustrative rubric. Four falsifiable propositions and a feasible pilot design were formulated for classroom evaluation. SPARC was not presented as an empirically validated intervention, and the available studies did not establish that GenAI necessarily caused durable declines in critical thinking. The paper concluded that educational evaluation should examine unaided transfer, reasoning quality and confidence calibration, rather than the quality of AI-assisted products alone.

Keywords: spatial cognition; generative artificial intelligence; cognitive offloading; scientific reasoning; critical thinking; higher education

1. Introduction

An AI-generated explanation may be fluent, convincing and wrong. In visually intensive disciplines this matters especially: a learner can accept an attractive representation without deciding where a force acts, how a structure changes when rotated, or whether an image agrees with physical constraints. The central issue is therefore not whether AI is used, but which cognitive operations remain with the student.

Evidence on this question is emerging but mixed. A survey of 319 knowledge workers found that greater confidence in GenAI was associated with less self-reported critical-thinking effort in AI-supported tasks; the finding is correlational and concerns workers rather than students (Lee *et al*., 2025). Among university students, AI dependence has been associated with lower self-reported critical thinking (Tian and Zhang, 2025), whereas GenAI use combined with feedback literacy and reflective thinking has been associated with stronger critical analytical skills (Yan *et al*., 2026). The clearest causal evidence comes from a field experiment with nearly one thousand secondary-school mathematics students: unrestricted GPT-4 access improved performance during practice, but students who had used it scored 17% lower than controls once access was removed, while a version with pedagogical guardrails largely avoided this harm (Bastani *et al*., 2025). Conversely, a randomised trial in university physics showed that an AI tutor built on research-based teaching principles more than doubled median learning gains relative to in-class active learning (Kestin *et al*., 2025). The effect of AI therefore appears to depend on instructional design.

Spatial-thinking research offers a complementary insight: constructing and interpreting visual representations is central to learning across STEM fields (Taylor *et al*., 2023), and spatial visualisation plays a measurable role in physics problem solving (Kozhevnikov *et al*., 2007). This paper asks: *How can AI-supported teaching protect students' spatial and scientific reasoning without rejecting the benefits of AI?*

2. Conceptual background

2.1 What connects spatial reasoning and critical thinking? Spatial reasoning includes imagining rotations and changes of perspective, understanding locations and relations, and using external representations to infer change (Taylor *et al*., 2023; Uttal *et al*., 2013). Scientific reasoning involves coordinating models with evidence, interpreting covariation and testing explanations. Critical thinking adds the appraisal of assumptions, evidence, alternatives and conclusions. These constructs overlap in tasks that demand representation-based inference, but they are not identical: an excellent mental-rotation score does not, by itself, establish strong argument evaluation.

Consider a student interpreting the velocity-time graph of a ball tossed vertically upward. Spatial processing is needed to map a one-dimensional trajectory onto a graph whose horizontal axis is time rather than space; scientific reasoning links the constant slope of the graph to the constant gravitational acceleration; and critical evaluation tests whether a claim such as "the acceleration is zero at the top" is consistent with that slope. Students with lower spatial-visualisation ability tend to read such graphs as literal pictures of the path (Kozhevnikov *et al*., 2007). The intersection proposed here is therefore a set of observable practices, namely constructing representations, deriving predictions, evaluating discrepancies and justifying revisions, rather than a claim of one common underlying ability.

2.2 When does AI offloading become educationally problematic? Cognitive offloading means shifting mental demands onto the environment (Risko and Gilbert, 2016). It is not intrinsically harmful: a calculator can release working memory for conceptual analysis, and an AI-generated alternative may reveal assumptions a learner missed. The concern arises when an external tool performs the very operation the task is intended to develop. This paper distinguishes *supportive offloading*, in which AI supplies examples, feedback or counterarguments after a learner's own attempt, from *substitutive offloading*, in which AI supplies the initial mental model, the inference and the final justification. The distinction is instructional and testable; it is not a diagnosis of any student.

Several recent findings motivate caution without justifying alarm. In a sample of 580 Chinese university students, AI dependence was associated with lower critical thinking, partly through cognitive fatigue (Tian and Zhang, 2025). Lee *et al*. (2025) found that GenAI shifted the locus of critical effort toward verification, response integration and task stewardship. The Bastani *et al*. (2025) experiment illustrates the substitutive pattern directly: students used unrestricted AI as a crutch, gained during practice and lost on unaided assessment. Conversely, a 2026 conceptual review argues that managed offloading can support higher-order learning (Marshall, 2026), and a survey of 421 undergraduates found that the positive association between GenAI use and critical analytical skills ran largely through GenAI feedback literacy, moderated by reflective thinking (Yan *et al*., 2026). Together these findings do not support a universal claim that "AI makes students think less"; they support the narrower claim that the sequence and purpose of AI use matter.

2.3 Why put spatial representations first? External sketches, diagrams and embodied transformations make reasoning visible to both student and instructor. Spatial skills are malleable: a meta-analysis of 217 training studies reported an average effect of g = 0.47 relative to controls, with transfer to untrained spatial tasks (Uttal *et al*., 2013). This does not establish transfer to critical thinking or demonstrate any GenAI effect. However, drawing a representation before consulting AI exploits the benefits of effortful generation, a well-documented desirable difficulty (Bjork and Bjork, 2011), and creates a baseline artefact that shows what the student produced independently. Such a baseline is especially valuable when an AI-produced diagram looks plausible but contains physical inconsistencies. Recent design principles for critical-thinking scaffolding similarly recommend preserving cognitive friction, treating language models as provisional thinking partners and alternating AI-free with AI-mediated phases (Vendrell and Johnston, 2026); SPARC operationalises these principles for spatially demanding tasks.

Putting spatial representations first is, however, not trivial for teachers. A narrative review of barriers and enablers of spatial teaching showed that its implementation is complex at the levels of curriculum, pedagogy, teachers' pedagogical content knowledge and everyday classroom practice (Bufasi *et al*., 2024). Hands-on activities, varied representations, alternative forms of assessment and professional learning were identified as enablers, whereas limited instructional time, insufficient teacher preparation, a weak link between cognitive research and classroom practice and difficulty in grading spatial work were identified as barriers. SPARC responds to this complexity by embedding spatial work in ordinary problem solving rather than adding separate spatial training, by keeping each artefact brief and by providing an explicit scoring rubric (Section 5.3).

3. Method and scope of the conceptual synthesis

This is a theory-building and pedagogical design paper in which sources were purposively selected to establish connections across four areas: (i) spatial thinking and its malleability; (ii) cognitive offloading and metacognition; (iii) empirical associations between generative AI (GenAI) use and critical thinking; and (iv) pedagogical scaffolds that position AI-generated outputs as objects of critical evaluation. Foundational reviews and recent peer-reviewed publications were prioritised. Because no exhaustive search, formal screening or quality weighting was undertaken, claims about prevalence and pooled effectiveness are deliberately avoided.

The design process followed three steps: identify the cognitive activity at risk when a model gives a complete answer; specify an observable student action that retains that activity; and formulate testable outcomes. SPARC is a proposed instructional artefact derived from these steps, not a model estimated from data.

4. The SPARC framework

SPARC is a five-stage sequence to be repeated for spatially or causally demanding problems (Figure 1, Table 1). Its key rule is temporal: learners make a traceable prediction before viewing any AI-proposed solution. AI is then used as a critic and a source of contrasting representations, not as an unquestioned answer key, and it is withdrawn again for the final transfer check.

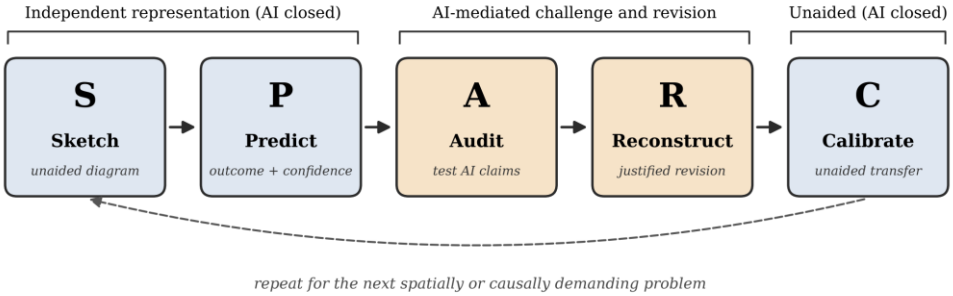


Figure 1. The SPARC sequence. Sketch and Predict build an independent representation with AI closed; Audit and Reconstruct use AI as a fallible comparison partner; Calibrate checks unaided transfer.

Table 1. Stages of the SPARC framework

Stage	Student action	Role of GenAI	Evidence retained
S Sketch	Draw, map or arrange the relevant quantities and relations unaided	None; no answer generation yet	Initial labelled diagram
P Predict	State an expected outcome and its causal reason	None; no solution revealed	Prediction and confidence (0-100%)
A Audit	Interrogate the AI output for assumptions, errors and rival interpretations	Generates an alternative, a critique or a deliberately uncertain explanation	Marked discrepancies with the principle or source used to check them
R Reconstruct	Revise the representation and reasoning; explain what changed	Offers hints or counterexamples, not a replacement final answer	Revised diagram and change log
C Calibrate	Solve an analogous item without AI and reassess certainty	Unavailable	Unaided answer and confidence

4.1 Mechanism and teacher controls: Sketch and Predict protect the generation phase. Audit exposes learners to alternative explanations and requires checking against physical constraints, principles or sources. Reconstruct demands an explicit account of conceptual change, and Calibrate separates apparent task success from independent understanding, the distinction that Bastani *et al*. (2025) showed to be decisive. Instructors should specify which AI functions are permitted at each stage, prohibit uploading identifiable personal or student data, and avoid grading the aesthetics of drawings. A useful audit prompt is: "Identify two physically plausible alternatives, one assumption that could invalidate this interpretation, and a test that would distinguish them. Do not give the final answer." Constraining the AI to hints rather than solutions mirrors the guardrails that made AI tutoring effective in controlled studies (Bastani *et al*., 2025; Kestin *et al*., 2025).

5. Worked teaching applications

5.1 Kinematics: the vertical motion of a tossed ball: Students toss a ball vertically upward and catch it at the release height. Before using AI, each student draws a motion diagram with velocity and acceleration vectors at five instants (release, halfway up, the top, halfway down and the catch), sketches predicted position-time, velocity-time and acceleration-time graphs, and states with a confidence rating what the velocity and the acceleration are at the highest point. Taking upward as positive and neglecting air resistance, the motion is described by

$$v = v_0 - g\,t, \quad t_{\text{top}} = v_0 / g, \quad h_{\max} = v_0^2 / (2g), \tag{1}$$

where v_0 is the release speed and $g = 9.81$ m/s^2. For $v_0 = 4.9$ m/s, Equation (1) gives $t_{\text{top}} \approx 0.50$ s and $h_{\max} \approx 1.2$ m. In the Audit stage the instructor presents an AI-style explanation containing a classic, plausible error: that both velocity and acceleration are zero at the top because the ball "stops". Students test the claim against their own vector diagram and against the straight, constant-slope velocity-time graph, and explain why a ball with zero acceleration at the top would remain there. They then reconstruct their graphs, for example correcting an acceleration-time graph that changed sign at the top or a position-time graph drawn as a picture of the path. In the Calibrate stage, students sketch the three graphs for a new case without AI, such as a ball thrown downward from a balcony or a cart rolling up and back down an inclined ramp. The vignette targets the velocity-acceleration distinction and the graph-as-picture error, both of which depend on transforming between spatial representations.

5.2 Introductory physics: work and power on a staircase: A student climbs a staircase at a constant average speed. Before opening AI, learners draw a side view, mark the vertical displacement h, define the system and predict whether the work done against gravity depends on the time taken. The AI is then asked to produce two explanations, one of which conflates work and power. Learners audit both using

$$W = m\,g\,h \quad \text{and} \quad P = W / t, \tag{2}$$

revise a labelled energy-flow sketch and, in the Calibrate stage, compare two climbing times without AI. For a 60 kg student rising 3.0 m, $W \approx 1.8$ kJ irrespective of time, whereas the mean power is about 290 W for a 6 s climb and about 150 W for a 12 s climb. The vignette deliberately uses a familiar setting to distinguish correct formula substitution from conceptual discrimination. Its purpose is not to "catch AI hallucinations" for entertainment but to make the learner state which physical principle falsifies an appealing answer. Instructor-designed erroneous examples should be disclosed in the debriefing so that students do not confuse a designed contradiction with the typical reliability of a particular model.

5.3 Assessing reasoning without rewarding AI fluency: Table 2 proposes indicators that score the student's reasoning trail rather than the polish of the final product.

Table 2. Illustrative rubric for SPARC artefacts

Indicator	Operational scoring suggestion (0-2)
Representation	0 absent or incoherent; 1 partly correct relations; 2 coherent, labelled spatial or causal structure
Prediction and warrant	0 unsupported; 1 plausible with weak warrant; 2 constraint-based prediction
Audit quality	0 accepts or rejects by authority; 1 identifies a discrepancy; 2 tests it against evidence or principle
Reconstruction	0 copied AI text; 1 partial revision; 2 justified conceptual revision
Transfer and calibration	Unaided transfer item scored separately; confidence-accuracy gap recorded

The rubric is illustrative and not yet validated; in a study, two independent raters and inter-rater agreement estimates are advisable. AI interaction logs can describe the process, but they should not substitute for evidence of student reasoning or be taken as proof of learning.

6. Research propositions and evaluation design

Four propositions follow from the framework, in line with recent calls to separate the efficiency of AI-assisted products from cognitive depth and delayed transfer (Lin and Al-Hada, 2026). *P1:* Compared with immediate access to AI answers, SPARC will produce higher-quality independent representations and justifications on equivalent tasks. *P2:* SPARC will yield better delayed, unaided transfer than unrestricted AI use, even if both groups produce similarly polished AI-supported products. *P3:* Learners using SPARC will identify more substantive AI errors and unsupported assumptions than peers who receive only a generic "verify the AI" warning. *P4:* Any benefit will be moderated by prior spatial skill, disciplinary knowledge, task complexity and AI feedback literacy rather than being uniform across learners.

6.1 Feasible pilot study: A practical pilot could use introductory physics sections, comparing SPARC with unrestricted AI support under equivalent task duration, prompts and instructional content. Random allocation of students within sections is preferable; where it is not feasible, or where contamination between conditions is likely, intact sections can be assigned and the study analysed as quasi-experimental. A third, AI-free comparison would help separate the effect of AI access from that of structured prompting if resources allow. Pre-test measures could include a standard spatial task such as the Purdue Spatial Visualization Test: Rotations (Guay, 1977), domain-concept questions and baseline confidence calibration. During the intervention, researchers would collect initial and revised sketches, predictions, audit notes, consented chat logs and time on task. Immediate and delayed post-tests would consist of parallel problems completed without AI. Ethical approval and informed consent are required, particularly for the collection of chat logs.
The primary outcome should be a preregistered score on delayed, unaided reasoning or transfer, scored blind to condition where practical. Secondary outcomes include detection of planted AI errors, representation quality, conceptual accuracy and the absolute gap between confidence and correctness. Analyses should adjust for baseline performance and nesting within teaching groups, and sample size and minimum detectable effects should be established before recruitment.

6.2 Risks and boundary conditions: SPARC increases task time, one of the main barriers to spatial teaching identified by Bufasi *et al*. (2024), and may impose unnecessary drawing demands on some learners; instructors are also likely to need professional learning to score spatial artefacts consistently. Not every problem requires a spatial representation; for non-spatial tasks a concept map or evidence diagram may be more appropriate. Students with low prior knowledge may need worked examples or prompts before they can predict independently, and accessibility alternatives should allow verbal, tactile or digital representations. Research should monitor whether the intervention merely increases reported effort without improving retention, and whether instructor-authored erroneous AI responses create avoidable confusion. The framework does not claim that spatial skills cause critical-thinking competence, that AI use necessarily erodes cognition, or that a single session produces lasting gains; its claim is that learner-authored reasoning can be preserved and its benefits tested.

7. Conclusions

A central problem in AI-supported education is the gap between an excellent product and evidence that the learner can reproduce the reasoning behind it. Spatial tasks make this gap especially visible, because diagrams can hide assumptions about orientation, dimensions, causal direction and measurement. SPARC responds by preserving an independent first representation and requiring a falsifiable prediction, a principled audit, a justified reconstruction and an unaided transfer check. It integrates spatial and critical reasoning as classroom practices without conflating them as psychological constructs.

The practical contribution is a lightweight, adaptable protocol for physics, engineering and other representation-rich subjects, which reframes AI from a default solution generator into a source of competing claims that students must test. The theoretical contribution is a distinction between supportive and substitutive offloading, together with conditional, falsifiable predictions. The empirical contribution remains prospective: controlled studies are needed to establish whether the protocol improves learning, for whom and at what cost. The most defensible educational target is not less AI use as such, but stronger evidence of what students can still explain, check and transfer when AI is absent.

References

[1]. Bastani H., Bastani O., Sungu A., Ge H., Kabakcı Ö. and Mariman R., 2025. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, Vol.122, No.26, e2422633122. https://doi.org/10.1073/pnas.2422633122

[2]. Bjork E.L. and Bjork R.A., 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In: Gernsbacher M.A., Pew R.W., Hough L.M. and Pomerantz J.R. (Eds.), *Psychology and the Real World*, Worth Publishers, New York, pp.56-64.

[3]. Bufasi E., Lin T.J., Benedicic U., Westerhof M., Mishra R., Namsone D., Dudareva I., Sorby S., Gumaelius L., Klapwijk R.M., Spandaw J., Bowe B., O'Kane C., Duffy G., Pagkratidou M., Buckley J. *et al*., 2024. Addressing the complexity of spatial teaching: A narrative review of barriers and enablers. *Frontiers in Education*, Vol.9, 1306189. https://doi.org/10.3389/feduc.2024.1306189

[4]. Guay R.B., 1977. *Purdue Spatial Visualization Test: Rotations*. Purdue Research Foundation, West Lafayette, IN.

[5]. Kestin G., Miller K., Klales A., Milbourne T. and Ponti G., 2025. AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, Vol.15, 17458. https://doi.org/10.1038/s41598-025-97652-6

[6]. Kozhevnikov M., Motes M.A. and Hegarty M., 2007. Spatial visualization in physics problem solving. *Cognitive Science*, Vol.31, No.4, pp.549-579. https://doi.org/10.1080/15326900701399897

[7]. Lee H.-P., Sarkar A., Tankelevitch L., Drosos I., Rintel S., Banks R. and Wilson N., 2025. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ACM, Article 1121. https://doi.org/10.1145/3706598.3713778

[8]. Lin J. and Al-Hada N.M., 2026. The critical-thinking paradox in generative AI-integrated learning: Distinguishing efficiency from cognitive depth. A differentiated framework and testable propositions. *Frontiers in Psychology*, Vol.17, 1906070. https://doi.org/10.3389/fpsyg.2026.1906070

[9]. Marshall J., 2026. Conceptually demystifying the intersection between higher education students' GenAI use and critical thinking. *Discover Education*, Vol.5, 399. https://doi.org/10.1007/s44217-026-01392-2

[10]. Risko E.F. and Gilbert S.J., 2016. Cognitive offloading. *Trends in Cognitive Sciences*, Vol.20, No.9, pp.676-688. https://doi.org/10.1016/j.tics.2016.07.002

[11]. Taylor H.A., Burte H. and Renshaw K.T., 2023. Connecting spatial thinking to STEM learning through visualizations. *Nature Reviews Psychology*, Vol.2, pp.637-653. https://doi.org/10.1038/s44159-023-00224-6

[12]. Tian J. and Zhang R., 2025. Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta Psychologica*, Vol.260, 105725. https://doi.org/10.1016/j.actpsy.2025.105725

[13]. Uttal D.H., Meadow N.G., Tipton E., Hand L.L., Alden A.R., Warren C. and Newcombe N.S., 2013. The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, Vol.139, No.2, pp.352-402. https://doi.org/10.1037/a0028446

[14]. Vendrell M. and Johnston S.-K., 2026. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Computers and Education: Artificial Intelligence*, Vol.10, 100572. https://doi.org/10.1016/j.caeai.2026.100572

[15]. Yan Z., Fang Z. and Zou W., 2026. The association between generative AI use and university students' critical thinking: A moderated mediation model of GenAI feedback literacy and reflective thinking. *Frontiers in Psychology*, Vol.17, 1939286. https://doi.org/10.3389/fpsyg.2026.1939286