International Journal of Social Science and Education Research Volume 9 Issue 11, 2026
ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

Reforming Engineering Mechanics Teaching through Generative AI: A Model-Evidence-Verification Approach

Yuan Chang

School of Civil Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215011, China

Abstract

An answer to an engineering mechanics problem can read fluently and still rest on the wrong body, support, or boundary condition. The weakness shows up plainly once students turn to generative artificial intelligence (AI), which produces convincing derivations even when the underlying mechanical model is shaky. The teaching design proposed here asks students to commit to their own model first, obtain a critique or a competing interpretation from AI, test the claims against mechanics evidence, revise the model if the evidence demands it, and finally solve a related problem with no help from AI. The checks are ordinary disciplinary tools: free-body diagrams, equilibrium, compatibility, units, limiting cases, and an independent calculation. A truss lesson illustrates the sequence in action when students cannot agree about a roller reaction or about when a zero-force rule applies. The rubric that accompanies the design scores the model, the evidence brought to bear on it, the stated reason for revision, and the transfer performance. The proposal is offered as a workable teaching arrangement and as a starting point for classroom evaluation.

Keywords

Engineering mechanics; generative artificial intelligence; model construction; verification; formative assessment; transfer.

1. Introduction

In most Engineering Mechanics exercises the hard decisions are made before the student ever starts. The body is named, the supports are drawn, the loads placed, the required quantity stated. That format serves calculation practice well, but it leaves little room for the choices that matter in real analysis—what to isolate, and what an idealized support can actually transmit. A student may thus produce the expected number on the strength of a poor free-body diagram or an unjustified member model [1-5]. No amount of correct algebra repairs that.

Generative AI makes the weakness impossible to ignore. A polished solution appears within seconds, yet it may drop a reaction, treat a two-force member as an ordinary rigid body, or build on a boundary condition the diagram never licensed. Tests in mechanical engineering indicate that the models perform unevenly, weakening on graphical, spatial, and multistep tasks as the problem form changes [6,7]. The real danger in class is not a wrong answer; it is that a well-written answer can pass as evidence.

The same problem points to a constructive use of AI in teaching. An instructor can present its response as a claim to be examined rather than a solution to copy. Students might, for instance, compare two choices of system boundary or judge whether a stated roller reaction is even possible. The free-body diagram and the governing equations decide the dispute; the fluency of the response decides nothing.

This idea is developed here as a five-stage sequence for an ordinary Engineering Mechanics course. The student models first, consults AI second, verifies what the tool claims, records any

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

revision, and closes with a related problem solved without AI. The paper ties the sequence to established work on mechanics learning, shows how it maps onto common course topics, and walks through a truss lesson in some detail. A rubric and a deliberately modest plan for classroom testing close the paper.

2. Literature and Design Basis

Introductory mechanics research has returned again and again to representational and conceptual difficulties that numerical accuracy alone does not capture. The Statics Concept Inventory grew out of an attempt to measure persistent misconceptions about forces, connections, and equilibrium [1]. In reinventing statics instruction, the emphasis has correspondingly fallen on conceptual understanding, active engagement, and the link between physical situations and formal analysis [2]. Studies of free-body diagrams add that students often fail to use these representations productively, even when they can reproduce familiar procedures [3-5]. Together these results argue for treating model construction as a learning outcome in its own right, not a preliminary sketch that precedes calculation.

Broader accounts of engineering design and expertise point the same way. Judgment grows out of coordinating representations, assumptions, analysis, and iteration, not out of picking formulas [8,9]. Work on model-eliciting activities likewise links problem formulation to the assessment of open-ended engineering solutions [10]. For mechanics teaching, the implication is that students need situations in which several idealizations look plausible at once, and must then defend why one of them suits the stated purpose. The defense matters because an idealization is neither simply true nor false; it is useful under particular loading, geometric, material, and accuracy conditions.

Assessment has to follow the same objective. Constructive alignment insists that learning activities and assessment criteria reward precisely the reasoning a course claims to value [11]. Formative feedback works best when learners can hold their current work against a goal and act on the difference [12]. Intelligent tutoring research offers a parallel caution: timely guidance supports learning, but the assistance has to be shaped around the cognitive activity the task targets [13]. Within this framework, AI accordingly enters only after an initial student model exists, and its output is treated as material for judgment, never as a substitute for it.

Recent writing on generative AI in education sets out both promise and peril. On the one hand, quick feedback, multiple explanations, and adaptable examples; on the other, hallucination, opacity, bias, overreliance, and pressure on assessment validity [14,15]. Engineering education studies add a practical warning: a conventional take-home problem can be solved by AI without ever demonstrating student understanding [6]. These concerns translate into four design requirements. The first attempt must be unaided; accepting or rejecting an AI claim must rest on visible evidence; revision quality, not prompt sophistication, is what gets assessed; and the course must end with an unaided transfer task.

3. Model-Evidence-Verification Teaching Reform

3.1. Framework and Learning Sequence

Modeling in engineering mechanics is a string of choices that a finished calculation easily conceals: which body to isolate, which interactions to draw, how to idealize a connection, which relation governs. Verification asks whether those choices hold up under checks drawn from mechanics. Matching an answer key counts as one such check, and usually not the most informative one.

The sequence opens with a brief, unaided attempt. Each student names the body or subsystem, marks the external actions, states the main idealizations, predicts the response, and writes

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

down the governing relations. The record does not need to be polished; its only job is to capture the student's first model before an AI response can anchor it.

Students then turn to AI—or to a common response the instructor has prepared—for a criticism or a competing interpretation. A genuinely useful response questions the chosen boundary, flags a missing assumption, or argues that some member is really a two-force member. At this stage the response carries no special authority. It is just another model, and it must be checked like any other.

Verification pulls the discussion back to mechanics. Does every external interaction appear on the free-body diagram? Are the constraints and the load path physically possible? Do equilibrium, compatibility, and the constitutive relations actually support the claim? Units, signs, symmetry, limiting cases, and a second calculation tend to expose errors quickly. A numerical model can add evidence, but its output by itself says nothing about whether its modeling assumptions are appropriate; experimental observations are what test those assumptions, within the conditions studied. Afterward students record what they changed and why—or, if the original model survived the checks, why it did. Either way, the record ties a decision to specific evidence rather than treating agreement with AI as improvement.

The cycle closes with a related problem worked entirely without AI. Changing a support, a load direction, the geometry, or the target quantity is normally enough to stop simple copying while keeping the underlying concept recognizable. The task shows whether the student can rebuild and check the model on their own. Fig. 1 summarizes the complete cycle. Every transition demands a visible student artifact, which keeps the instructional focus on mechanical judgment rather than prompt production.



Figure 1. Five-stage Model-Evidence-Verification teaching cycle

3.2. AI Roles and Governance

AI use stays bounded. Students name the tool and retain the prompt and response used for the task; they submit no personal or confidential material; and the instructor watches for misconceptions that could feed unsafe engineering judgments. Responsibility for the submitted model rests with the student.

4. Course-Level Implementation

The course materials for Engineering Mechanics give the design a concrete home. They run through introductory modeling, rigid-body statics, equilibrium and friction, trusses, deformable-body analysis, material behavior, torsion, and beam bending, and they describe a 32-hour course in which coursework carries 40% and the final examination 60% of the grade. Within that existing schedule, a set of carefully chosen worked examples can carry the reform without disturbing the normal course units.

Four checkpoints suit an initial trial: support reactions and free-body diagrams, friction state, truss member modeling, and beam internal-force diagrams. Each demands a different kind of decision, yet all rely on methods already in the slides. A checkpoint takes over part of the usual worked-example discussion. The class sees the problem diagram without its solution, then a single common AI response, then a record that ties each disputed claim to mechanics evidence; only after that is the original worked solution shown.

Fifteen minutes per checkpoint is a workable starting allocation: three for the individual model, two to read the AI response, five for paired verification, two for the revision record, three for a

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

short no-AI transfer item. The split is a proposal, to be adjusted once a trial has run. The instructor prepares and checks the AI response beforehand; if a response is edited to plant a teaching error, it must be labelled an instructor-adapted example and never reported as unaltered model output.

Shorter verification tasks cover the remaining topics. The deformable-body slides make the connection between equilibrium, geometric compatibility, and material relations explicit, which helps students see why equilibrium by itself cannot solve a statically indeterminate problem. Material curves give evidence for choosing a constitutive model. Torsion exercises separate torque, stress, and rotation; beam exercises open independent checks through section equilibrium, the load-shear-moment relations, and boundary conditions. Table 1 summarizes the four checkpoints.

Table 1. Initial teaching checkpoints using existing course slides

Course topic	Modeling focus	AI-supported activity	Mechanics evidence
Support reactions	Body isolation and constraints	Check a disputed reaction	Support motion; free-body diagram; equilibrium
Friction equilibrium	Direction and limiting state	Check an assumed friction limit	Motion tendency; friction inequality; equilibrium
Plane trusses	Two-force members and zero-force rules	Check a member-force claim	Joint equilibrium; independent section
Beam internal forces	Sign conventions and load changes	Check a shear or moment diagram	Left/right cuts; diagram jumps; boundary values

4.1. Illustrative Truss Lesson

The plane-truss materials already cover the method of joints and the method of sections, so a worked-example diagram from them can serve as the common problem. Three extra slides are prepared: the diagram with an unaided modeling question, the AI claim to be checked, and a short evidence record; the original solution stays available for discussion. For the transfer item, a related review problem is adapted by altering a load or support orientation, with a check that the revised structure remains stable.

The illustrative lesson uses a pin-jointed roof truss under vertical joint loads. A pin supports one end; an ideal frictionless roller on a horizontal surface supports the other; and somewhere in the arrangement sits a joint where a zero-force-member rule looks tempting. The objectives are to produce a defensible whole-truss free-body diagram, assign member-force directions under a consistent tension assumption, say when a zero-force rule actually holds, and confirm a result by an independent method.

At the start each student gets a drawing and a brief design context, not a finished analytical model. Working alone, they fix the system boundary, mark the external reactions, state the idealizations, predict the sign of selected member forces, and opt for either the method of joints or the method of sections. The instructor collects or photographs these first attempts before any AI interaction. A two-minute peer explanation may follow, but the record is not yet edited.

The AI phase runs on a prepared prompt or a shared instructor-generated response: 'Review this truss model. Identify assumptions that must hold, propose one alternative interpretation of the supports or members, and state which conclusion is least certain.' Everyone receives the same response, precisely so the activity does not turn into a hunt for the best prompt. The response is treated as a set of claims by design. One claim may have the frictionless roller on the horizontal surface supplying a horizontal reaction; another may have an unloaded joint

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

automatically making both connected members zero-force members. Detached from the diagram, both can sound reasonable.

Groups then fill in an evidence sheet. Each claim must be matched with a support symbol, a free-body diagram, an equilibrium relation, or a structural condition. They check the whole-truss reactions, isolate the disputed joint, and set the method of joints against a strategically chosen section. A sign discrepancy prompts a re-examination of the assumed member-force direction, not a blanket rejection of the calculation. Symmetry, when the geometry allows it, serves as a prediction that is then checked against equilibrium—never as a stand-in for it.

Next comes a revision record from each student. A worked example: the original model put two reaction components at the roller; the AI response repeated that reading; the support symbol and the permitted direction of motion showed that only the vertical reaction, normal to the horizontal surface, belongs there; the revised free-body diagram drops the unsupported horizontal component; the equilibrium equations are rebuilt. A student who got the support right from the start may instead document why the model was kept and which evidence supported that choice. Success is therefore not counted in changes made.

The lesson closes with an individual transfer problem solved without AI. The support orientation differs, one load is inclined, and the target member has moved. Students must rebuild the free-body diagram and supply one independent check—a task short enough to finish in class, yet different enough that replaying the earlier sequence will not work. The responses tell the instructor whether more instruction is needed before the class moves on to frames and machines.

5. Assessment and Prospective Evaluation

The rubric looks at five things: completeness of the model, quality of its assumptions, the mechanics evidence used, the reason offered for revision, and performance on the no-AI task. A complete model carries the boundary, forces, reactions, dimensions, and unknowns the analysis requires. Strong evidence might be a corrected diagram, an equilibrium check, compatibility, units, or a second method of solution. Prompt length and agreement with the AI response earn no credit on their own.

Table 2. Analytic assessment rubric for student artifacts

Dimension	4 Exemplary	3 Competent	2 Developing	1 Initial
Model completeness	All essential model elements included	Minor noncritical omission	Key element missing; model still interpretable	Boundary or interactions incorrect
Assumption quality	Explicit, appropriate, condition-linked	Appropriate but partly implicit	Some assumptions unjustified	Absent or inconsistent assumptions
Mechanics evidence	Several relevant checks support judgment	One correct mechanics check	Partial or weakly linked evidence	Relies on AI output or answer matching
Revision reasoning	Decisions fully traceable to evidence	Most revisions explained	Revision recorded; reasoning limited	Acceptance or rejection unexplained
No-AI transfer	New model independently built and verified	Correct core model; minor error	Partial transfer with substantial support	Cannot build a defensible model independently

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

A classroom trial ought to answer a handful of practical questions. Do students produce better models after running through the cycle? Which checks survive when AI is taken away? What evidence moves them to accept, qualify, or reject a machine-generated claim? The first attempt, the revision record, the transfer response, the rubric score, and a short reflection give an initial study its material. Concept-inventory items can be added where they match the course outcomes [1].

A stronger test would use comparable sections or alternating course units, with prior achievement and baseline modeling recorded. A second rater should score a sample so the rubric's interpretation can be verified. Numerical results should be read next to the revision records, which show how each change came about. Before any data collection, a formal study must settle consent, institutional review, privacy, and the storage of AI interaction logs.

Reporting calls for details: the AI tool, its version, the prompt conditions, and the account arrangement. When the research question is about the teaching sequence itself, a single common response for the whole class removes one avoidable source of variation. The exact response used in class deserves an archive, since model behavior drifts over time.

Each checkpoint leaves behind a small paper trail: the initial free-body diagram or model, a claim-and-evidence sheet, a revision record, and a no-AI response. The rubric in Table 2 applies directly to this set. In a pilot, the records can steer feedback inside the existing coursework component; altering grading weights would have to follow the course assessment arrangements. Final-examination items should keep demanding independent analysis. Setting assisted work beside a later unaided item distinguishes finishing the classroom activity from actually retaining the modeling method.

6. Discussion

The design casts AI in a role other than tutor. Its job is to supply a plausible claim that students must judge. Authority stays with the support symbol, the free-body diagram, and the mechanics relations. In that arrangement, AI literacy is learned inside the ordinary work of the discipline, not in a separate lesson about prompts.

The design is not free. An unaided attempt consumes class time, a shared AI response limits personalization, and revision records add marking. Keeping checkpoints short and sampling selectively holds the workload in check. For some topics an instructor-written contrasting case works just as well, especially when access is uneven or the likely AI response would add little to the lesson.

None of this has yet been tested against student outcomes or transfer data. The truss lesson is a single bounded example; dynamics, laboratory work, and advanced mechanics may call for different checks. Results will also ride on the particular AI system and institutional rules in force at the time. A classroom study should accordingly start with a few topics, record teacher workload and student access, and hold the mechanics standard fixed while varying the form of the challenge.

7. Conclusion

Generative AI throws a familiar weakness of Engineering Mechanics teaching into relief: a finished derivation can conceal a poor model. The sequence proposed here converts that weakness into a teaching problem. Students commit to a model, meet a competing claim, test it against mechanics, explain whatever they revise, and finally work without AI. The question is never whether the machine produced the right answer; it is whether the student can defend the model. The next step is to run the sequence in ordinary course units and see what survives on the transfer task.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

References

[1] Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. Journal of Engineering Education, 94(4), 363–371. https://doi.org/10.1002/j.2168-9830.2005.tb00864.x.

[2] Steif, P. S., & Dollar, A. (2005). Reinventing the teaching of statics. International Journal of Engineering Education, 21(4), 723–729.

[3] Rosengrant, D., Van Heuvelen, A., & Etkina, E. (2009). Do students use and understand free-body diagrams? Physical Review Special Topics - Physics Education Research, 5, 010108. https://doi.org/10.1103/PhysRevSTPER.5.010108

[4] Savinainen, A., Mäkynen, A., Nieminen, P., & Viiri, J. (2013). Does using a visual-representation tool foster students' ability to identify forces and construct free-body diagrams? Physical Review Special Topics - Physics Education Research, 9, 010104. https://doi.org/10.1103/PhysRevSTPER.9.010104

[5] Aviani, I., Erceg, N., & Mešić, V. (2015). Drawing and using free body diagrams: Why it may be better not to decompose forces. Physical Review Special Topics - Physics Education Research, 11, 020137. https://doi.org/10.1103/PhysRevSTPER.11.020137

[6] Nikolic, S., et al. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. European Journal of Engineering Education, 48(4), 559–614. https://doi.org/10.1080/03043797.2023.2213169

[7] Frenkel, M. E., & Emara, H. (2024). ChatGPT-3.5 and -4.0 and mechanical engineering: Examining performance on the FE mechanical engineering and undergraduate exams. Computer Applications in Engineering Education, 32(6), e22781. https://doi.org/10.1002/cae.22781

[8] Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. Journal of Engineering Education, 94(1), 103–120. https://doi.org/10.1002/j.2168-9830.2005.tb00832.x

[9] Litzinger, T. A., Lattuca, L. R., Hadgraft, R. G., & Newstetter, W. C. (2011). Engineering education and the development of expertise. Journal of Engineering Education, 100(1), 123–150. https://doi.org/10.1002/j.2168-9830.2011.tb00006.x

[10] Diefes-Dux, H. A., & Salim, W. W. A. W. (2012). Transforming the first-year engineering experience through authentic problem-solving: Taking a models and modeling perspective. Procedia - Social and Behavioral Sciences, 56, 314–332. https://doi.org/10.1016/j.sbspro.2012.09.660

[11] Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364. https://doi.org/10.1007/BF00138871

[12] Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

[13] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

[14] Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202611_9(11).0006

[15] Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535