DOI:10.5937/IIZS25060T

MODELS OF AI INTEGRATION IN ENGINEERING EDUCATION: A PRACTICAL FRAMEWORK FOR LEARNING OUTCOMES, RETENTION, AND ACADEMIC INTEGRITY

Nemanja Tasic[1], Igor Vecstejn[1], Vuk Amizic[1], Tamara Milic[1]
[1]University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
e-mail: nemanja.tasic@tfzr.rs, igor.vecstejn@tfzr.rs, vuk.amizic@tfzr.rs, tamara.milic@tfzr.rs

Abstract: This paper proposes a practical, program-level framework for integrating generative AI into engineering education while safeguarding academic integrity and improving learning outcomes and retention. Drawing on recent evidence from engineering and STEM education, the framework aligns curriculum outcomes with ABET expectations, embeds AI-supported active learning (e.g., retrieval practice, automated feedback), and redesigns assessment to prioritize process evidence and oral/interactive defenses. We describe governance mechanisms, recommended tool chains, and analytics for continuous improvement. A comparative table and implementation roadmap illustrate how departments can sequence adoption across courses and years of study. The approach is intended for resource-constrained programs seeking measurable, ethically grounded gains in performance, persistence, and fairness.
Key words: engineering education, generative AI, learning outcomes, retention, academic integrity, assessment design

INTRODUCTION

Engineering programs operate under simultaneous pressures: rapid AI capability growth, industry demand for AI-literate graduates, and external accountability via accreditation. As programs redesign learning environments, three tensions recur: (I) how to maintain rigorous conceptual understanding when assistance tools shorten problem-solving paths; (II) how to preserve fairness and academic integrity without an unproductive arms race with detection; and (III) how to deliver measurable gains in outcomes and persistence that justify institutional investment [1].
This paper is positioned as a pragmatic blueprint rather than a technology showcase. We delineate governance choices (what is permitted, expected, and assessed), instructional patterns (how AI augments-not replaces-core practices), and measurement (how retention, attainment, and integrity are tracked over time) with exemplars aligned to ABET outcomes and contemporary guidance from national and international bodies [2].
Generative AI is moving from pilot use to routine practice across higher education, including engineering programs. Institutions now face pressure to equip students with AI fluency while ensuring mastery of core principles and ethical practice [1]. At the same time, accreditation bodies require programs to demonstrate attainment of program outcomes and effective assessment practices, which must be reconciled with the realities of AI-enabled learning [1].

BACKGROUND AND RELATED WORK

A complementary meta-analysis across 51 (quasi)experimental studies reports large effects on performance and moderate gains in perception and higher-order thinking, and highlights moderators such as learning model and role of ChatGPT, underscoring the need for scaffolded design [3].
Recent meta-analytic evidence indicates positive average effects of ChatGPT on academic performance and higher-order thinking, with substantial heterogeneity across course type and duration; this supports measured, context-sensitive adoption rather than blanket policies [4].

Syntheses across STEM and engineering education show heterogeneous effects: AI excels at code generation, structure for lab reports, and formative hints, but can hallucinate sources, mis-handle units, and propagate subtle misconceptions in modeling tasks. Consequently, scholars recommend structured tool alignment (match task → tool → evidence), human-in-the-loop verification, and explicit instruction on prompt design and result checking [5].

On integrity, converging evidence warns that empirical evaluations show that AI-text AI-text detectors have documented limitations (precision/recall trade-offs and false positives) (e.g., false positives) and can be inequitable, especially for non-native writers; consequently, design-based integrity strategies are preferred (higher false positives for L2 writers), and that design-based approaches process evidence, oral defenses, and authentic tasks—outperform surveillance-heavy strategies in both fidelity and student trust [6,7]. illustrative policy frameworks (e.g., sector bodies and ministries) converge on disclosure norms and instructor-designed safeguards rather than blanket bans [6,7].

Recent reviews in engineering education report rapid growth of AI-powered tools for coding assistance, simulation, lab report drafting, and formative feedback. Studies comparing common assistants (e.g., ChatGPT, Copilot, Gemini, Wolfram) find differential strengths across task types, suggesting the need for structured tool alignment to course outcomes [6].

Evidence from STEM learning science indicates that retrieval practice with spacing can increase durable learning, and emerging work shows that generative AI can scale the creation of retrieval prompts and adaptive practice sequences [2].

Concurrently, multiple universities and national bodies caution against reliance on AI detectors and instead recommend assessment redesign and explicit policy guidance to uphold academic integrity in the presence of generative models [8].

FRAMEWORK OVERVIEW

The framework is intentionally lightweight. Each component has two deliverables: (A) a one-page playbook for instructors and (B) a minimal dataset to observe change. Outcomes alignment yields a mapping table and a short list of assessable artifacts; policy yields course-level AI statements and a disclosure template; AI-enhanced teaching yields ready to use prompts and workflows; assessment design yields rubrics emphasizing process and oral defense; analytics yields a two-page dashboard with 6–8 indicators refreshed each term [1].

Implementation is iterative: choose two courses and one lab for a pilot; collect baseline metrics; then scale via communities of practice. The objective is not tool maximalism but dependable, auditable gains in learning and retention with integrity safeguards embedded by design. [8]

We propose a five-component framework that operates at the program-level: (1) outcomes alignment; (2) policy and ethics; (3) AI-enhanced teaching; (4) assessment design with process evidence; and (5) analytics for continuous improvement. Fig. 1 presents the architecture linking curriculum, policy, teaching, assessment, and analytics.

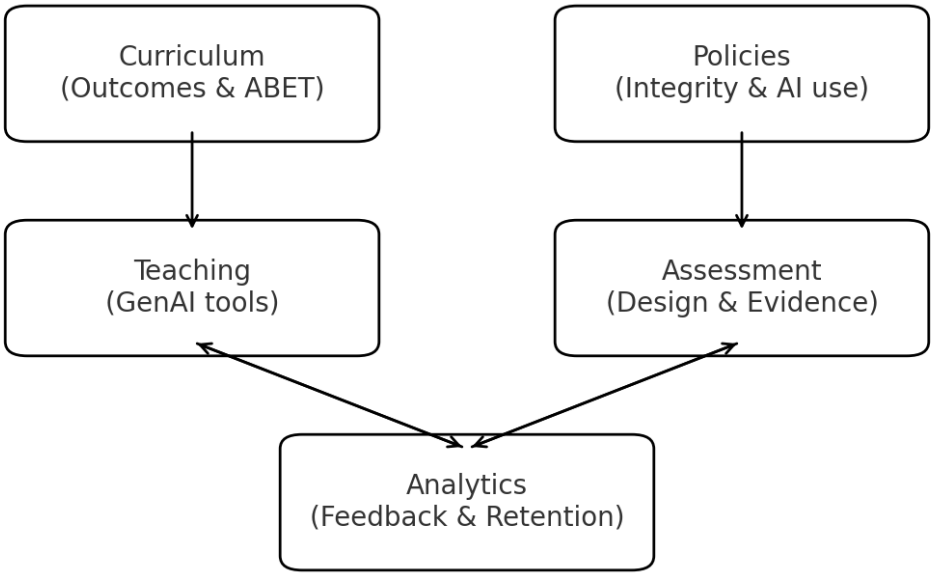

Fig. 1. *Program-level framework for AI integration in engineering education.*

OUTCOMES ALIGNMENT

For each course outcome, instructors specify the contribution of AI: ideation aid, scaffolding, or feedback generator. They also specify what remains strictly human-assessed (e.g., oral synthesis, troubleshooting under constraints). Performance indicators are phrased to make AI use observable: "documents alternative designs generated with and without AI, justifies selection via constraints," or "explains verification steps for AI-produced analyses" [9]. (see Table 1).

Table 1. *Example mapping of AI-enabled activities to ABET outcomes and assessable evidence. [9]*

ABET outcome (EAC 2025-2026, abridged)	AI-enabled activity	Evidence artifact
1. Solve complex engineering problems	AI-augmented ideation and trade-off analysis	Design log with prompt history & alternatives
3. Communicate effectively	AI-assisted drafting with self-explanations	Drafts with tracked changes + reflection
6. Experimentation	AI for data cleaning & exploratory analysis	Lab notebook, code, and validation plan
7. Acquire & apply new knowledge	AI-curated retrieval practice & reading plans	Practice logs and performance trend
Ethics & responsibility	Disclosure of AI use; model risk analysis	Signed AI use statement; error analysis

Alignment also encompasses equity: AI-generated scaffolds (worked examples, question prompts) are provided to lower the floor without lowering the ceiling. Students document use through a concise disclosure form incorporated into submissions. These artifacts feed assessment and analytics pipelines.

Programs should map AI-enabled activities to established student outcomes (e.g., ABET Criterion 3) such as problem solving, experimentation, communication, and ethical responsibilities. AI use should explicitly support, not replace, performance indicators tied to these outcomes [9].

Example mappings include: AI-augmented design ideation to broaden alternatives, AI-supported data analysis to improve rigor, and structured prompting for technical communication practice, each with human oversight.

POLICY AND ETHICS

Course AI statements include: (1) permitted uses by task type; (2) required disclosure format; (3) responsibilities for accuracy, privacy, and citation; and (4) consequences for misrepresentation. Because detector-only regimes are brittle, policies emphasize teachable integrity: students are assessed on how they used, verified, and learned from AI, not whether they avoided it entirely [8].

Ethics instruction is integrated into labs and projects (model limitations, bias, and error propagation). Students practice model risk analysis on their own prompts and produce short error-analysis notes demonstrating verification literacy. This cultivates habits aligned with professional codes and ABET expectations [9].

Departments should publish course-level AI use statements that specify permitted and prohibited uses, citation requirements, and accountability for errors. Policies should discourage over-reliance on automated detectors and emphasize teachable integrity—for example, by requiring disclosure of AI assistance and including reflective justification of prompts and revisions [8].

AI-ENHANCED TEACHING

Three patterns have proven classroom-ready: (i) Retrieval practice at scale—AI generates varied items aligned to outcomes; instructors vet seeds, then analytics schedule spacing; (ii) Draft and critique—students iterate with AI on reports or design memos while maintaining a change log; (iii) Coding and simulation hints—AI suggests tests and refactoring plans, with students validating and benchmarking performance [2].

To mitigate overreliance, instructors cap the proportion of AI-scaffolded work per assignment, require a reflection on the delta between initial and final understanding, and apply the rubric's verification dimension. Early adopters report time savings in feedback cycles that are reinvested into higher order coaching [1].

Instructors can combine generative AI with evidence-based strategies. Practical options include automated feedback on early drafts, hint generation for code and simulations, and AI-generated retrieval practice cards with spaced scheduling. These supports should be calibrated to avoid cognitive overload while keeping students in the loop [1]. (see Table 2)

***Table 2**. Assessment redesign tactics and the integrity risks they mitigate.*

Assessment tactic	Integrity risk mitigated	Implementation note
Oral defense/viva	Contract cheating; AI-only drafting	Short 5–10 min per team; rubric for conceptual grasp
Process portfolio (code/commit history)	Undeclared AI assistance; ghostwriting	Collect prompts, drafts, and diffs as evidence
In-lab practical with variable data	Answer sharing; memorized solutions	Randomize parameters; require reasoning steps
Studio critique/design review	Hallucinated claims; shallow understanding	Peer + instructor questions; require citations
Open-AI take-home with disclosure	Detector over-reliance; hidden use	Award marks for the quality of AI use and verification

ASSESSMENT DESIGN WITH INTEGRITY

We move from detection to design. Studio style reviews and vivas surface genuine understanding. Process portfolios preserve prompt histories, intermediate outputs, and decision rationales; graders evaluate reasoning, crosschecks, and improvements across versions. In lab practicals, use parametrized datasets or novel parts so that memorized or fully outsourced solutions confer little advantage [10].

Marking emphasizes verification: what authoritative sources were used; what unit checks, boundary tests, or sanity checks were performed; how disagreements between AI outputs and references were resolved. Where appropriate, students perform brief oral demonstrations (2–5 minutes) to defend choices under probing questions. [6]. See Table 3.

Table 3. *Rubric for evaluating AI-assisted coursework using process evidence.*

Dimension	Exceeds expectations	Meets expectations	Approaches expectations	Required evidence
Problem framing & ethics	Explicit risk analysis of AI use; citations/checks for claims	States allowed uses; acknowledges limitations	Vague policy compliance; misses key risks	AI use statement; prompt log
Method & iteration	Multiple prompt/draft iterations with justified changes	One iteration with a brief rationale	Minimal iteration; unclear rationale	Version history; diffs
Accuracy & verification	Triangulates outputs with authoritative sources and tests	Performs basic checks; a few sources	Little verification; relies on AI verbatim	Verification checklist; test results
Communication	Clear, well-structured technical writing with figures/tables	Understandable with minor issues	Disorganized; missing visuals	Final report; presentation
Academic integrity	Consistent process evidence; oral defense demonstrates mastery	Mostly consistent; can answer core questions	Gaps in process or defense; partial mastery	Oral defense notes; assessor rubric

Given the limitations of detection, assessment should be redesigned to make unauthorized outsourcing less beneficial and to value authentic performance. Tactics include oral defenses, studio critiques, process portfolios (with commit histories), and practical labs with variable data. Research and field reports indicate that AI-generated submissions can bypass current detectors, reinforcing a shift toward assessment-design strategies and even score well, underscoring the need for design-based integrity [6]. (see Table 4)

Table 4. *Program implementation roadmap with measurable KPIs*

Term	Key actions	Primary KPIs (targets)
Semester 1 (Pilot)	Adopt AI-supported retrieval practice in 2 core courses; publish course AI-use statements; staff training	Quiz to exam alignment ≥0.6; ++8–12 percentage-point retention gains on spaced items [6]; 100% disclosure compliance
Semester 2 (Scale)	Extend to labs; implement process portfolios & oral defenses; baseline outcome mapping to ABET [1]	Outcome attainment +5 pp; integrity incidents ≤ baseline; student satisfaction ≥ 4/5

Semester 3–4 (Institutionalize)	Governance for tool selection/privacy; annual review of outcomes; dashboarding & equity monitoring	DFW rate −10%; equity gap −5 pp; Replace reliance on detectors with design-based academic-integrity approaches (authentic tasks, process evidence, oral defenses) [7]

ANALYTICS FOR CONTINUOUS IMPROVEMENT

A minimal indicator set prevents data sprawl: (I) outcome attainment (rubric based); (II) item level retention on spaced materials; (III) integrity incidents (categorized by root cause); (IV) equity gaps by demographic proxy measures where permissible; (V) student experience (brief pulse); and (VI) workload (grading time) [1].
Triangulation matters: LMS traces alone are insufficient. We combine process artifacts (prompt logs, diffs), rubric scores, and small-n instructor notes each term to interpret anomalies and adjust scaffolds or policies. Dashboards show trends rather than league tables to support learning-oriented improvement. [1]
Programs should track a small set of indicators: course-level learning outcomes achievement, retention across terms, assessment integrity incidents, and equity gaps. Classroom analytics—from LMS logs to AI tool usage telemetry—can inform rapid cycles of improvement when triangulated with instructor judgment and student feedback [1].

IMPLEMENTATION ROADMAP

Pilot governance: the department nominates a lead, defines a change window (one term), selects courses, and allocates support (TA hours, consultation). Baselines are recorded for attainment, DFW, and integrity. Success is pre-registered as directional improvement against baselines, not absolute targets in the first term [1].
Scaling prioritizes coherence over coverage: materials and rubrics are templated; communities of practice share failures as well as wins. Procurement and privacy teams review tools for data handling; fallback plans are documented to avoid single vendor lock-in. [9]
Phase 1 (pilot): identify two core courses to pilot AI-enhanced retrieval practice and feedback. Establish disclosure policies and train staff on tool capabilities and limitations [8].
Phase 2 (scale): extend to lab courses with process portfolios and oral defenses; align mappings to program outcomes and collect baseline indicators [9].
Phase 3 (institutionalization): incorporate governance for tool selection, privacy, and data security; review outcome attainment annually; iterate on assessments and analytics dashboards [8]. (see Table 5)

Table 5. *Comparison of traditional and AI-integrated delivery across key dimensions.*

Dimension	Traditional Delivery	AI-Integrated Delivery
Learning Outcomes	Manual feedback, limited iterations	Automated feedback, rapid iteration, adaptive tasks
Retention	Cramming, low retrieval scheduling	Spaced retrieval, AI-curated practice
Academic Integrity	Proctoring-centric	Assessment redesign, process evidence, oral defense
Equity & Access	One-size-fits-all	Personalized scaffolding and supports
Workload	Instructor-heavy grading	AI-assisted grading with human moderation

To illustrate potential retention gains, Fig. 2 shows a simulated trajectory comparing a control condition with an AI-supported spaced retrieval approach, consistent with reported effects in STEM contexts [2].

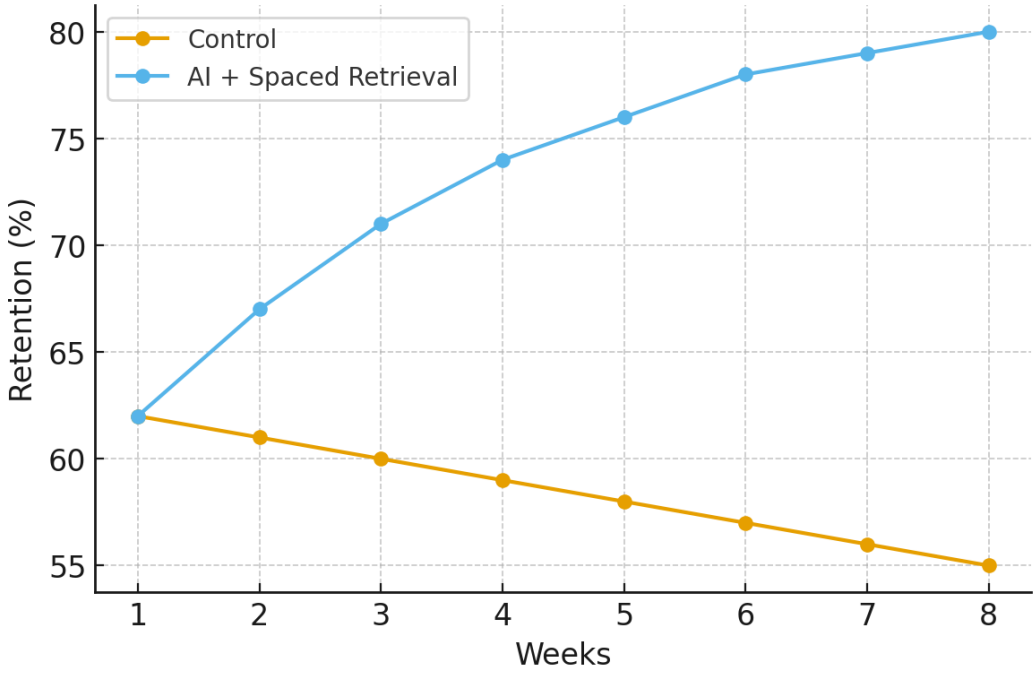

Fig. 2. *Illustrative retention gains with AI-supported spaced retrieval. [2]*

DISCUSSION

Our approach balances innovation with accountability. By tying AI use to specific outcomes and assessable artifacts, we avoid both lax permissiveness and inflexible prohibition. The primary risk is uneven instructor capacity; we mitigate via ready to adopt playbooks, shared banks of vetted prompts, and lightweight calibration sessions. [1]

Another risk is policy drift: without periodic review, allowances may expand without safeguards. Annual audits against ABET criteria and institutional policies ensure practices remain justifiable and student centred. The strongest evidence of success will be sustained improvements in retention on conceptually demanding topics with narrowing equity gaps [2].

The framework's novelty lies in its program-level alignment and assessment redesign grounded in integrity-by-design, rather than detection. While tooling evolves quickly, anchoring practices in established outcomes and learning science helps curricula that are resilient to rapid AI change [1].

Limitations include variability in tool performance across tasks and disciplines, staff capacity for change management, and the need to ensure accessibility. Early institutional pilots highlight both promise and risks, reinforcing the importance of governance and staff development [9].

CONCLUSION

Generative AI can be a leverage point for better learning, not a shortcut around it when embedded in outcomes, instruction, and assessment with integrity by design. The framework, instruments, and KPIs offered here are intentionally simple so that departments can execute quickly and iterate responsibly. Future work should test the framework in multi institution pilots and report comparative cost benefit and equity impacts using shared indicators [1]. Engineering programs can responsibly integrate generative AI by aligning

activities to outcomes, clarifying ethical use, redesigning assessment, and using analytics to improve learning and retention. The proposed framework offers a practical path to measurable gains while safeguarding academic integrity.
Additional implementation tables complement the framework and support adoption choices across courses. To operationalize the framework, the roadmap and rubric below provide concrete milestones, KPIs, and evaluation criteria.

REFERENCES

[1] U.S. Department of Education, Office of Educational Technology (2023). Artificial Intelligence and the Future of Teaching and Learning. URL: https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf
[2] Bego, C. R., et al. (2024). Single-paper meta-analyses of the effects of spaced retrieval practice in STEM. International Journal of STEM Education.
[3] Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanities and Social Sciences Communications, 12, 621
[4] Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224
[5] Baig, M. I., Yadegaridehkordi, E., et al. (2024). ChatGPT in higher education: A systematic literature review. Computers & Education: X
[6] Christianson, J. S. (2024). End the AI detection arms race. Patterns, 5(10), 101058
[7] Liang, P., et al. (2023). GPT detectors are biased against non-native English writers. Patterns. (PMCID: PMC10382961)
[8] UNESCO (2023, updated 2025). Guidance for generative AI in Education and Research. URL: https://unesdoc.unesco.org/ark:/48223/pf0000386693
[9] ABET. Criteria for Accrediting Engineering Programs, 2025–2026
[10] Jisc National Centre for AI (2024). Embracing generative AI in Assessments: A Guided Approach (Flowchart)
[11] JISC - Embracing GenerativeAI in Assessments:A Guided Approach. URL: https://nationalcentreforai.jiscinvolve.org/wp/files/2024/08/Embracing-Generative-AI-in-Assessments-Flowchart.pdf