What Would Education Look Like With an AI Literacy Program Based on Proven Methods?

Basil C. Puglisi, MPA
A Human-AI Collaboration
basilpuglisi.com

In a 2026 randomized trial at a large public university, 2,379 students across 30 instructors received access to an AI tutor built into their courses. Final grades fell by 0.37 standard deviations, and participation in course activities fell by 0.90 standard deviations (Liu et al., 2026). Most instructors left the tutor in its default direct-instruction mode, and students used it without assignments or guided exercises. First-generation students lost more than twice as much as continuing-generation students.

At Harvard, researchers built a different AI tutor on seven practices from learning research, among them active learning, managing cognitive load, scaffolding, and timely feedback. In a crossover trial of 194 students, that tutor produced more than double the median learning gains of an in-class active learning lesson (Kestin et al., 2025).

The same kind of tool produced opposite results, and what changed was the method around it. The method in the second trial came from education's own research. The OECD's Digital Education Outlook 2026 warns that generative AI "may enhance the apparent quality of student work . . . without improving their actual learning" (OECD, 2026, p. 20).

Education does not need to invent a new pedagogy for AI, because its own research has already tested the methods that work. It needs one element those methods never had to supply: a named human who owns the decision at a checkpoint. What follows describes an AI literacy program built that way, from a high school classroom to a professional certificate.

The methods education already owns

Active learning is the clearest case. Prince (2004) defines it as any instructional method that engages students in the learning process, so that they do meaningful activities and think about what they are doing. A meta-analysis of 225 studies in undergraduate science, engineering, and mathematics found that active learning raised examination scores by about 6 percent (Freeman et al., 2014). Students taught by traditional lecture were 1.5 times as likely to fail.

Scaffolding is a second case. Support builds capability when it matches the learner's reach and is then withdrawn, so the learner takes on what it once supplied (Vygotsky, 1978; Wood et al., 1976). Interactive and constructive engagement produces more durable learning than passive reception (Chi & Wylie, 2014). Metacognition, the monitoring of one's own thinking, names the capacity a learner uses to question an answer before accepting it (Flavell, 1979).

Research on AI points the same way. Cognitive offloading helps when it frees the mind for the hard part of a task and harms when it replaces the thinking (Risko & Gilbert, 2016). Designs that make people think before they accept AI output reduced overreliance on AI in decision-making (Buçinca et al., 2021). In a field experiment with nearly 1,000 high school mathematics students, unstructured AI access raised practice grades and then lowered exam grades by 17 percent once access ended. A version designed to scaffold rather than answer largely mitigated that loss (Bastani et al., 2025).

None of these studies tests the program described below, because none placed a named human with binding authority at a checkpoint and measured the result. Together they support its premises: structure decides the outcome, polished work is not evidence of learning, and reviewing is not the same as deciding.

Where the current approach falls short

Education measures the wrong thing because it does not yet know what to look for. The university in the 2026 trial watched use through a console of anonymized prompts and weekly usage emails, which counts activity rather than learning (Liu et al., 2026). Its instructors were not prepared to choose the pedagogy, so the default setting chose it for them.

Frameworks have not closed that gap. The AILit framework from the OECD and the European Commission names the right competences (European Commission & OECD, 2026). It then houses the checkpoint inside the learner's own self-monitoring and leaves every step optional (Puglisi, 2026a). Earlier critiques reached the same root. Institutions deployed AI without the method standards they require of every other practice, then blamed the tool for what unstructured use produced (Puglisi, 2026c).

What the program would teach

The program would open with a class called Ethical AI: The Ground Rules, and no assignment would use AI until that class was complete. The class treats AI as an environment rather than a single tool, reaching from search engines and grammar checkers to recommendation systems, agents, and robots. Its first rule asks whether something should be done before asking whether it can be done.

The first AI-assisted assignment then produces the learner's first artifact. Each learner writes three to five challenge prompts for the learner's own work and uses them on one real task. The learner then records what the prompts revealed that would have been missed without them, so the governance habit forms before the technical learning starts (Puglisi, 2025).

From there, every AI-assisted session follows one sequence, called the Checkpoint Session. In Human First, the learner writes a position, a plan, or a prediction, along with the measure that would show it worked, before any AI responds. Direct frames the task, assigns roles, and designates the sources that count, and Execute runs the work. Challenge then tests the output against a distinct AI system, a primary source, a peer, or a test. Verify traces each claim to a source outside the model, since a model that argues against its own answer still produces generated text rather than a calibrated measure. Own closes the checkpoint as accept, modify, or reject, with the rationale recorded and the learner's signature.

Before signing, the learner answers three questions: "Can I explain this decision to a colleague? Do I know the sources? Would I bet my reputation on this?" (Puglisi, 2025). The answers enter the checkpoint log beside the decision, so a reviewer can see what the learner weighed before owning the work.

Checkpoints come in proportion, because a teacher who signed every prompt of every student would produce ceremony rather than oversight. The learner signs routine and reversible work. The teacher signs milestones, graded work, escalations, and a weekly sample of learner checkpoints, while the institution signs tool adoption, privacy, and assessment design. A learner whose sample shows passive acceptance moves up a tier until the pattern breaks.

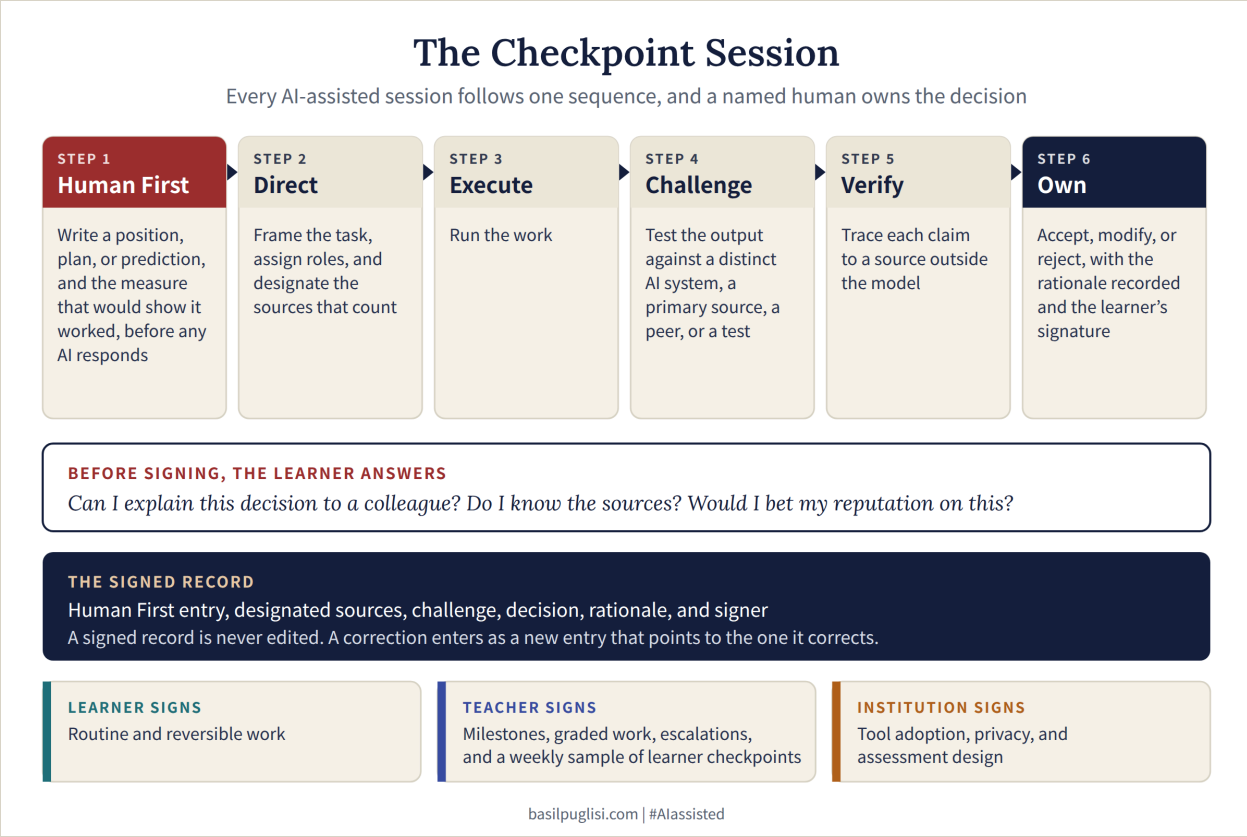


Figure 1. The Checkpoint Session: six steps, three questions before signing, and a signed record that is never edited.

Support is withdrawn on a published schedule. Early courses supply the checkpoint templates, the role assignments, and the approved sources, and later courses require learners to write their own. Each capstone then asks the learner to design a checkpoint system that another person would operate.

What the program would measure

A program built on method has to show its outcome, including what remains without the tool. Each learner would be measured with the Human Enhancement Quotient before the program begins, at the end of each term, and upon completion. The instrument is designed to read governance competence in a person (Puglisi, 2026b). Each term would also close with a check that removes the tools and measures retention and transfer to an unfamiliar problem. A learner who scores well with the tools and poorly without them may be showing the offloading pattern the program exists to prevent.

The capstone ends in a two-part defense. In the first part, with no AI, the learner explains each checkpoint decision and finds errors an assessor has seeded into a copy of the learner's own work. The learner then solves an unfamiliar problem, still without AI. In the second part, the learner applies the full sequence to a new problem with the tools present.

Every AI-assisted decision leaves a signed record of the Human First entry, the designated sources, the challenge, the decision, the rationale, and the signer. A signed record is never edited, and a correction enters as a new entry that points to the one it corrects. Oversight evidence that can be rewritten after the fact shows only that governance was intended (Puglisi, 2025). The record carries no emotion inference and no biometric data, and it is deleted when its retention period ends.

The learner is the customer of every program, and a person needs employment from the learning. Each program would therefore follow its graduates into work and track three employer outcomes. Employer adoption counts the employers who host learners, return for more, and hire again. Hiring records whether a graduate is hired into the work the program prepares for. Employment records whether that graduate is still employed at three months, six months, and each year after. Those outcomes are read beside the checkpoint record, since an employer can value fast output from a graduate who rarely challenges what the AI produces.

Who teaches it

Teachers come first. In the 2026 trial, most instructors left the AI tutor at its default setting, and both grades and participation fell (Liu et al., 2026). No teacher or faculty member would lead a studio in this program before completing the same foundation the learners take and reaching the same measures. The instructor is the named human who holds the authority to modify or reject any deployment that fails to produce learning.

Three programs, one standard

The standard enters the education system where it already trains people for work. It follows New York's rules as the working example, since every state sets its own structure. Each program splits in two. Half is a shared foundation, and the other half is one track: Professional and Leadership, Technical and Engineering, or Teaching and Enablement.

- **Grades 11 and 12 through BOCES.** A two-year, half-day career and technical education program of 900 hours, with 120 hours of work-based learning.
- **The associate degree and the minor.** A 64-credit Associate in Applied Science whose 18-credit core also anchors a minor open to any major.
- **The professional certificate.** Eighteen months at 8 hours a week for people who already hold a degree, where the degree becomes the domain of practice.

A partner may lower costs by working with companies and drawing on existing courses, such as industry credentials and openly licensed course materials. Each one would still run inside the Checkpoint Session and its record. The full design, with courses, hours, credits, and approval routes for all three programs, is on the AI Literacy Programs page.

What remains unproven

Checkpoint-Based Governance is in production use in the author's own work and has not yet been independently validated or deployed at enterprise scale (Puglisi, 2026d). The Human Enhancement Quotient is a working instrument without independent validation, and its assessor reliability has not been established (Puglisi,

2026b). The author also designed the frameworks these programs apply, which creates a conflict of interest in judging whether they work. The evaluation therefore belongs to people who did not design them.

The newest trials already supply two comparison conditions: a tutor built from learning research, and open access without required structure (Kestin et al., 2025; Liu et al., 2026). A program running the Checkpoint Session beside them would test what neither trial did, measuring retention, transfer, tool-removed performance, error detection, completion, and placement against comparison groups.

Practicing the standard

Critique earns its standing when the critic submits a design to the standard the critique set. This design asks education to do what its own research already supports. It puts the learner's thinking first, requires challenge before acceptance, withdraws support on schedule, and measures what remains when the tools are gone. A reader who finds a gap should hold the design to that same standard, and the AI Literacy Programs page sets out each program in enough detail to test it.

References

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122 (Correction: *Proceedings of the National Academy of Sciences, 122*(34), e2518204122. https://doi.org/10.1073/pnas.2518204122)

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

European Commission & OECD. (2026). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education*. OECD Publishing. https://doi.org/10.1787/65cd27d4-en

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Kestin, G., Miller, K., Klales, A., Milbourne, T., & 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*. https://doi.org/10.1038/s41598-025-97652-6

Liu, J., Sweet, T., Chen, M. H., Engelberg, J., Masters, M., Clark, M., Persaud, A., Lancaster, A., Hollingsworth, J., & Rice, J. K. (2026). *The effects of course-integrated AI tutoring on student performance and engagement: A randomized university trial* (EdWorkingPaper No. 26-1598). EdWorkingPapers. https://edworkingpapers.com/ai26-1598

OECD. (2026). *OECD digital education outlook 2026*. OECD Publishing. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Puglisi, B. C. (2025). *AI learning roadmaps for professionals*. basilpuglisi.com. https://basilpuglisi.com/ai-learning-roadmaps-for-professionals/

Puglisi, B. C. (2026a). *The continued failure in AI literacy: AILit produced a starting point halfway through the race and called theory a framework*. basilpuglisi.com. https://basilpuglisi.com/ailit-framework-rule-optional/

Puglisi, B. C. (2026b). *Measuring the governance competence at the center of AI literacy: The Human Enhancement Quotient (HEQ), the Augmented Intelligence Score (AIS): Connecting AI literacy and cognitive development through governance* [Working paper]. basilpuglisi.com. https://basilpuglisi.com/measuring-governance-of-ai-literacy/

Puglisi, B. C. (2026c). *Stop blaming AI for what the education system abandoned*. basilpuglisi.com. https://basilpuglisi.com/stop-blaming-ai-for-education/

Puglisi, B. C. (2026d). *What AI transformation actually requires*. basilpuglisi.com. https://basilpuglisi.com/ai-transformation-requirements/

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

The Other AI: Audio Briefings on Augmented Intelligence and AI Governance

Spotify | Apple Podcasts | Amazon Music | YouTube Playlist

#AIassisted using the HAIA Ecosystem | CC BY-NC-SA 4.0 Free for personal, educational, and noncommercial research use with attribution. Commercial exploitation, paid productization, and enterprise commercialization require separate permission and licensing.