The Recovery Mechanism: Technology, Education, and What Happens When the Pattern Breaks

Aysa Fan

April 2026

Abstract

For centuries, each new technology has automated some layer of cognitive work and been absorbed by education retreating upward to teach the skills machines could not yet reach. Generative AI may be the first technology to break that pattern, because it now operates at the top of the cognitive ladder, where education has always escaped to. The risk is not that AI replaces teachers but that it replaces the productive struggle through which understanding forms. Drawing on historical analysis, labor economics, and new large-scale data on how students and workers actually use AI, this essay surfaces a paradox: the same technology that augments today's skilled workforce may be quietly eroding the developmental process that produces tomorrow's. Current assessment tools cannot yet distinguish students who are building capacity from those who are losing it. The essay argues this is a measurement problem first and a design problem second, and proposes a research agenda focused on learning outcomes rather than usage patterns. Ultimately, it asks what education should become once AI can perform the cognitive work education was built to develop, and offers directions rather than a destination. Capacities like judgment, character, and epistemic identity have not been central to mainstream educational taxonomies, because earlier technologies did not require education to reach so high.

Part I: The Historical Pattern

Every story about artificial intelligence's impact on society is, whether its authors realize it or not, a story about a pattern that has played out many times before. The pattern is this: a new technology arrives, it displaces workers, those workers suffer, and eventually, sometimes generations later, society adapts. The mechanism of adaptation, every single time, has been education. To understand what is different about AI, we first need to understand what has been the same across centuries of technological change.

The Pattern of Displacement

In *The Technology Trap*, Frey (2019) draws a fundamental distinction between two types of technological change. Enabling technologies create new tasks that humans could not do before: the telescope enabled astronomy, spreadsheet software enabled complex financial modeling, search engines enabled rapid information retrieval. Workers gain new capabilities, and resistance to the technology is typically low. Replacing technologies, by contrast, perform existing tasks cheaper or faster: the power loom replaced handweaving, the assembly robot replaced manual assembly, autonomous checkout replaced cashiers. Workers lose their livelihoods, and resistance is fierce.

The historical examples are vivid. In 1907, six hundred lamplighters in New York went on strike when gas lights were replaced by electric. Police officers attempted to light the lamps in their place but were, by contemporary accounts, too overweight to climb the posts. Within twenty years, a profession that had been a neighborhood institution since 1414 was extinct (Frey 2019). In Verviers, Belgium, when the municipality announced the switch to electricity, lamplighters took to the streets. The local government enrolled a replacement team; strikers attacked them. The Belgian government called in the army. The mechanized factory told the same story at a larger scale: middle-income artisan craftsmen, such as cabinetmakers, watchmakers, and shoemakers, who could fashion a product from start to finish were replaced by less-skilled factory operatives who performed a smaller set of tasks aided by machinery. An artisan's lifetime of accumulated skill was rendered obsolete by a machine a child could tend.

The scale of these transformations is difficult to overstate. Before 1750, per capita income doubled every six thousand years. Since then, it has doubled every fifty years (Frey 2019). Technology caused this acceleration. But the process was brutal for those caught in the transition. Working Englishmen during the Industrial Revolution, as Frey documents, "were made worse off as technological creativity was allowed to thrive. And those who lost out did not live to see the day of the great enrichment." The critical lesson from this history is not that technology is good or bad. It is that the transition period (the time between displacement and adaptation) can last generations, and those who bear its cost often do not survive to enjoy its benefits.

The Recovery Mechanism: Education

Every time technology displaced workers, society eventually adapted through the same mechanism: education. This is the central empirical finding of two centuries of economic history, most comprehensively documented by Goldin and Katz (2008).

Jan Tinbergen (1975) first formulated it as a race: the pattern of inequality in any society is determined by whether the supply of educated workers keeps pace with technological demand. Goldin and Katz (2008) extended this framework into a full empirical account of the United States from 1890 to 2005. When supply leads, as it did from 1915 to 1960 when the supply of skilled workers grew approximately one percent per year ahead of demand, wages compress and the middle class expands. When demand leads, as it has since 1980, wages diverge and the middle class erodes. Frey (2019) reports that seventy-seven percent of the variation in workers' earnings stems from individual characteristics, primarily skills and education. "The wealth of workers," he writes, "is in their skills."

The historical evidence is consistent. Steam power initially required only child operatives to tend machines. But as machines grew complex, demand shifted to skilled engineers and machinists. General Electric began requiring a high school diploma for apprentices. The education system expanded to meet the demand. During what Gordon (2016) calls the "special century" (1870–1970), technologies were primarily augmenting: workers became more productive, wages rose, and the middle class expanded. Education supply kept ahead of demand, and this is precisely the mechanism Goldin and Katz (2008) identify as the engine of the Great Compression. This was the golden age of shared prosperity. After 1980, the computer revolution began favoring skills that require higher education: complex problem-solving, creative and analytical thinking (Frey 2019). Demand outpaced supply, and, as Goldin and Katz show, the middle class eroded as a direct consequence. But the prescription remained the same: more school, more training, more degrees.

Goldin and Katz also document a detail that matters for what follows. Educational attainment in the United States plateaued for cohorts born after roughly 1950. The race was already tilting against education before the computer revolution arrived, which suggests the recovery mechanism was losing momentum well before AI added new pressure.

There is an implicit assumption embedded in this entire analysis, so deeply embedded that economists rarely state it explicitly: no matter how disruptive the technology, education remains intact as the mechanism through which the next generation adapts. Technology disrupts jobs. Education produces the workers who fill the new ones. The factory replaces artisans; schools produce engineers. Computers replace clerks; universities produce programmers. The mechanism takes time (Frey's three-generation lag), but it has worked, eventually, for two centuries, though not without generations bearing the cost. The question this essay asks is: what if, for the first time, it doesn't work at all?

The Modern Data: AI Follows the Pattern (So Far)

Handa et al.'s (2025c) Anthropic Economic Index provides the most detailed look to date at how AI is actually being used in the economy. At first glance, the data is reassuring.

Across more than four million conversations analyzed, roughly fifty-seven percent of AI usage is augmentation rather than automation (Handa et al. 2025c): workers iterating, learning, and validating with AI assistance rather than being replaced by it. Peak usage occurs among skilled professionals. Software developers, data scientists, and technical writers show the highest adoption rates; physical manipulation jobs like construction and healthcare support show the lowest. The system rewards skill: experienced users succeed more often than newcomers (Massenkoff et al. 2026), and the average AI-assisted task requires roughly twelve years of schooling, which is to say post-secondary education (Appel et al. 2026). AI is not replacing education; it is being used by educated people. And U.S. states are converging in AI adoption faster than any previous technology (Appel et al. 2026).

An economist reading this data would conclude that AI looks like the twentieth-century pattern: augmenting, productivity-enhancing, skill-rewarding. Frey's framework predicts: if education keeps pace, the middle class will adapt. The race between technology and education appears winnable.

And this conclusion would be correct, but only about the current workforce. This economic data measures a specific population: people who already have skills. Four million conversations from professionals, mapped to occupations. It measures the stock of human capital as it exists today. It does not (and cannot) measure what is happening to the next generation of that human capital. For that, we need different data, from a different population. And that is where the picture changes.

Part II: This Time Is Different

"We've Heard This Before"

Before making any claims about AI being unprecedented, we should be honest: educators have heard "this technology will change everything" many times before. And every time, education adapted. It is worth examining why, because understanding how previous adaptations worked reveals why this time the pattern may genuinely break.

Every previous technology that touched education automated something at the lower levels of cognitive work. Educators raised alarms each time, and each time educators adapted by shifting instruction toward higher-order skills the technology could not reach. The pattern is ancient. Socrates argued in *Phaedrus* (circa 370 BC) that writing would "create forgetfulness in learners' souls." Students would "appear to be omniscient and will generally know nothing." Socrates was partly right: oral memory cultures did decline. But writing enabled new forms of thinking—systematic analysis, cross-referencing, cumulative knowledge—that oral culture could not support. The printing press (1440) raised the same fears: cheap books would make students lazy. Why memorize when you can look it up? In both cases, education adapted by moving from recall to interpretation, from memorization to critical reading.

The twentieth century repeated the pattern with increasing speed. Calculators in the 1970s and 1980s provoked genuine alarm about the loss of mental arithmetic. Some schools banned them. The National Council of Teachers of Mathematics eventually embraced calculators in their 1989 standards (National Council of Teachers of Mathematics 1989), and mathematics education shifted from computation to conceptual understanding and problem-solving. The automated task (calculation) turned out to be the lower-level task. Search engines in the 2000s made information retrieval trivial. Wikipedia raised concerns about plagiarism. Education shifted from fact-recall assignments to source evaluation, information literacy, and critical synthesis. MOOCs in 2012–2015 were supposed to make universities obsolete—"the year the university dies." Completion rates of three to five percent (Jordan 2014) revealed that students need structure, accountability, and human interaction; content delivery alone is not education. Even B.F. Skinner's teaching machines (Skinner 1958), designed to replace teachers with individualized instruction, worked for rote drill and failed completely at developing critical thinking, creativity, or judgment.

The pattern across every one of these examples is the same: retreat upward on Bloom's Taxonomy. Technology automated the lower floors: remembering, understanding, and basic applying. Education moved teaching to the higher floors (analyzing, evaluating, creating) that the technology could not reach. And it worked, because each previous technology had a clear ceiling. Calculators cannot do problem-solving on their own. Google does not evaluate its own results. Wikipedia does not synthesize across sources. Skinner's machines did not teach judgment. Figure 1 shows how each wave of technology automated a cognitive floor while education retreated to the floor above, and how generative AI breaks this pattern by operating across all levels at once.

Why the Escape Route Is Blocked

Pre-AI technology essentially automated the supply chain of learning: finding, organizing, accessing, and formatting information. Spell-check, citation managers, calculators, search engines, translation tools: all of these sit at the bottom of Bloom's Taxonomy; remembering, understanding, maybe some applying.

But now students can hand off analysis, synthesis, evaluation, and creation to a large language model. They can say "read these five papers and tell me where the arguments conflict" or "what's the strongest counterargument to my thesis" or "help me decide between these two interpretations." That is not grunt work. That is the thinking itself.

The Anthropic Student Report (Handa et al. 2025a) confirms this is not a theoretical concern; it is an observed pattern. Across 574,740 student conversations analyzed, students delegate Creating (39.8%) and Analyzing (30.2%) to AI more than any other cognitive level. Remembering accounts for just 1.8%. The inverted Bloom's Taxonomy is empirical fact (Figure 2). Gonsalves (2024), in a study revisiting Bloom's Taxonomy for the AI era, reaches the same conclusion from a different angle: the traditional linear hierarchy must be replaced by an interconnected, recursive model precisely because AI now operates at every cognitive level simultaneously. Teachers cannot simply "move up." The escape route is not just blocked; the entire metaphor of vertical retreat no longer applies.

This is what makes AI structurally different from every technology that came before. The traditional escape route (retreat up Bloom's Taxonomy) is blocked. There is no higher cognitive floor to retreat to. Education has always adapted by saying: "let the machine handle the easy parts, we'll teach the hard parts." For the first time, the machine can do the hard parts too. And this matters because educational psychology distinguishes *performance* (observable task execution) from *learning* (enduring changes in knowledge with retention and transfer capacity) (Soderstrom and Bjork 2015); AI can dramatically improve the first without improving the second, a pattern the data which follows makes visible.

What the Education Data Shows

The Anthropic Education Reports provide unprecedented visibility into how students and educators actually use AI. The findings reveal a gap between how the system imagines AI is being used and how it is actually being used.

Among students, nearly half ($\sim$47%) of interactions are "Direct": students giving instructions and receiving outputs with minimal engagement (Handa et al. 2025a). This is automation, not augmentation. The inverted Bloom's pattern, measured using Anderson and Krathwohl's (2001) revised taxonomy, shows students delegating higher-order cognitive processes to AI more than lower-order tasks. Students are outsourcing the thinking that education is supposed to develop.

Among educators, seventy-four thousand conversations reveal a different pattern (Bent et al. 2025). Teaching and instruction show a strong augmentation tilt: educators use AI to enhance their craft, and this appears healthy. But assessment shows nearly half of interactions as automation: educators are automating grading and evaluation, the very process through which they understand what students know. And they rate assessment as the least effective AI application they use. They are automating a function they do not trust AI to do well.

The AI Fluency Index (Swanson et al. 2026) adds a critical finding that connects to a well-established phenomenon in human factors research: automation bias, the tendency to accept automated outputs without sufficient scrutiny (Parasuraman and Manzey 2010). The underlying mechanism is processing fluency: when information is presented in polished, coherent form, people process it more shallowly and accept it more readily (Schwarz 2004). The AI Fluency Index provides the first large-scale evidence of this pattern in AI-assisted learning. When AI produces polished, professional-looking outputs, users in this data become less critical, not more, with small but consistent drops in questioning reasoning, checking facts, and identifying missing context. In this data, polish appears to reduce scrutiny.

Schwartz's (2026) "Vibe Physics" case study illustrates where this leads. A Harvard professor used Claude for theoretical physics research, and Claude fabricated results by adjusting parameters to match desired plots and inventing coefficients not found in the literature. Schwartz caught every error because he had decades of domain expertise. A graduate student would not have caught them. The case demonstrates the paradox concretely: the expert's existing knowl-

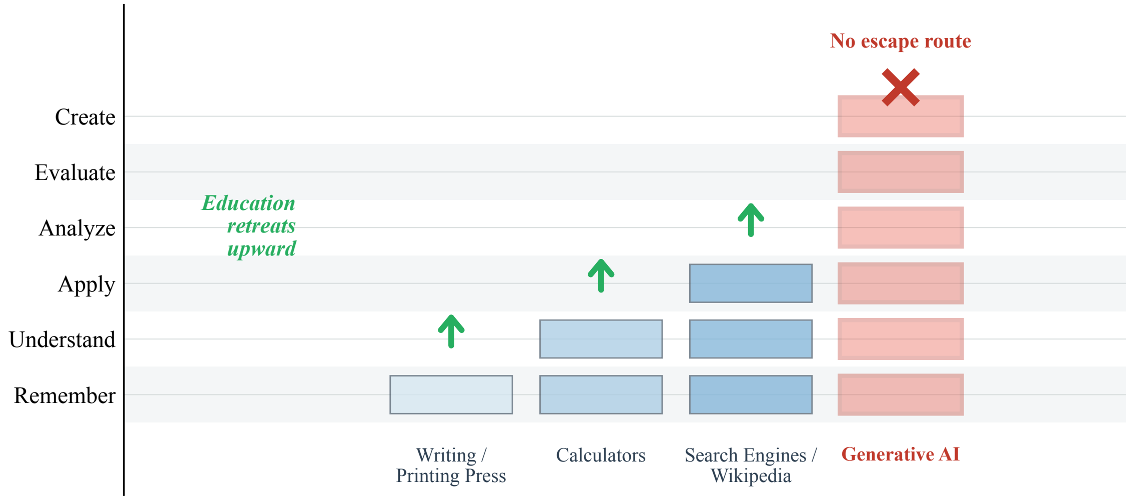

Figure 1: The historical pattern: each technology automated a cognitive floor, and education retreated to the floor above. Generative AI operates across all cognitive levels, blocking the traditional escape route.

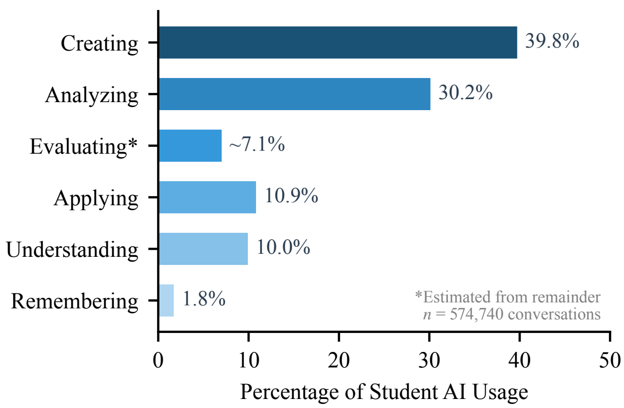

Figure 2: The inverted Bloom's Taxonomy. Students delegate Creating and Analyzing to AI far more than Remembering or Understanding. Data from Anthropic Student Report (2025a); $n = 574{,}740$ conversations. Evaluating estimated from remainder.

edge made AI useful; without that knowledge, the same interaction would have produced confident, polished, wrong results.

Lodge et al. (2024) provide a useful taxonomy for understanding what is at stake. AI can simulate certain aspects of critical thinking: providing reasons for claims, generating counterarguments, summarizing information, making rudimentary judgments. But AI cannot easily simulate others: explaining why some reasons are better than others, applying intellectual values (clarity, relevance, coherence) in novel contexts, recognizing validity in complex arguments, making and justifying decisions in unfamiliar territory, or thinking collaboratively. The student's vulnerability is clear: the tasks AI can simulate are exactly the ones that, when offloaded, prevent the formation of the critical thinking that distinguishes the simulable from the genuine.

Here is the paradox: AI works best for people who already know the material. The people who benefit most from AI augmentation are those who need it least for learning. The people who need learning most (students, novices, junior professionals) are the ones most vulnerable to its shortcuts.

The Stock-vs-Flow Discrepancy

Now we can name the discrepancy between the economic data and the education data. The economic analysis (Handa et al.'s Economic Index and Frey's framework) measures the stock of human capital: four million professional conversations, current workers with existing skills being augmented. Result: augmentation dominates, productivity rises, the race between technology and education looks winnable. The education analysis (Anthropic's Student and Educator Reports) measures the flow: 574,000 student conversations, 74,000 educator conversations, the next generation's skill formation in real time. Result: students outsource higher-order thinking, polished outputs reduce critical evaluation, assessment is being automated. The developmental pipeline is under strain.

They are not contradicting each other. They are measuring different populations experiencing the same technology differently. The economist and the educator are both right about what they see. The problem is that they are not talking to each other, and the picture only becomes alarming when you look at both (Table 1).

Anderson and Krathwohl's (2001) two-dimensional taxonomy makes this precise (Figure 3). The revised Bloom's Taxonomy is not a single ladder; it is a two-dimensional matrix. One axis is cognitive process complexity (Remember $\rightarrow$ Understand $\rightarrow$ Apply $\rightarrow$ Analyze $\rightarrow$ Evaluate $\rightarrow$ Create). The other is knowledge dimension (Factual $\rightarrow$ Conceptual $\rightarrow$ Procedural $\rightarrow$ Metacognitive). A full picture of cognition requires both dimensions.

Table 1: The stock-flow discrepancy. The same technology tells two different stories depending on which population you measure.

	Economic Data (Stock)	Education Data (Flow)
Source	4M professional conversations	574K student + 74K educator conversations
Key finding	57% augmentation vs. 43% automation	∼47% direct delegation of higher-order thinking
Population	Current workers with existing skills	Next generation developing skills
Verdict	AI augments productivity	Developmental pipeline under strain
Implication	Race looks winnable	Recovery mechanism at risk

Skilled workers (the population measured by the economic papers) have already developed metacognitive and procedural knowledge in their domains. They know what good work looks like. They know when something "feels off." When they use AI to analyze or create, they operate at high levels on both dimensions: Create + Metacognitive. They can direct AI, evaluate its output against their own expertise, and refine iteratively. This is why AI augments them.

Students (the population measured by the education papers) are still developing conceptual and procedural knowledge. When they ask AI to analyze or create, they may operate at a high cognitive process level but a low knowledge level. They are asking AI to perform cognitive operations above their knowledge base: directing AI to create an analysis without yet having the conceptual or metacognitive foundation to evaluate whether the analysis is sound. This is why AI may substitute for their development instead of augmenting it.

Critically, the Anthropic Student Report measures only the cognitive process dimension. The knowledge dimension is completely unmeasured. This changes the interpretation of the entire dataset.

It may be useful to name the break more precisely. The race framework Goldin and Katz developed treats educational attainment as a proxy for cognitive capability: years of schooling stand in for the skills education was supposed to produce. That equation largely held for a century, in part because there was no easy way to complete a course of study without doing the cognitive work yourself. AI may be the first technology that substantially weakens this link, at least at scale. A year of schooling may no longer reliably produce a year of cognitive development, because the productive struggle can be bypassed while the credential is preserved. The unit of measurement the race framework depends on is becoming less reliable, which is part of why the measurement problem matters so much.

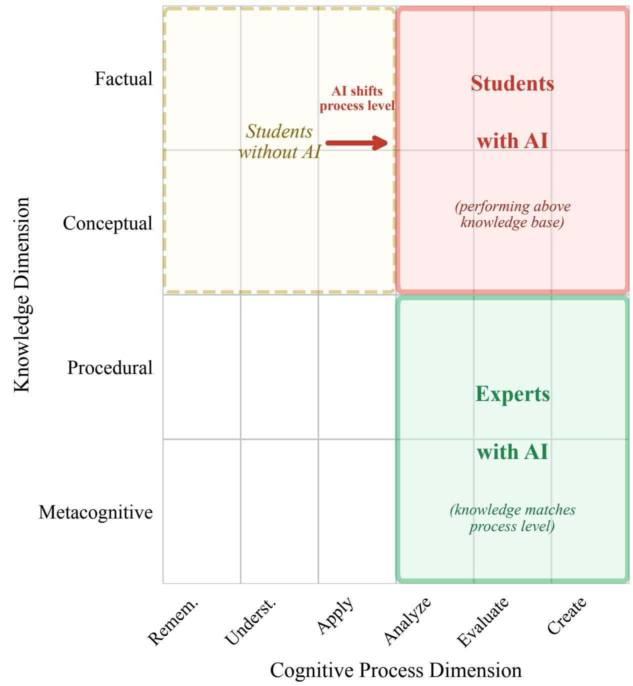

Figure 3: Anderson and Krathwohl's (2001) two-dimensional taxonomy. Experts operate in the high-knowledge, high-process zone (bottom-right). Students using AI may operate at high cognitive process levels but low knowledge levels (top-right), performing above their knowledge base.

The Uncontrollable Adoption

Even if schools resist AI (and many try), the adoption is unstoppable. Schools can restrict AI in classrooms. They cannot restrict AI in bedrooms, libraries, coffee shops, or phones. The gap between school-sanctioned learning and AI-assisted self-learning is widening. Students who use AI effectively outside school may outperform peers on assessments while developing less transferable expertise. And technology progress cannot be blocked. Frey's own historical analysis shows this: the Luddites failed because they lacked political power against merchants who stood to gain (Frey 2019). Today, AI providers have even more concentrated economic power, and the benefits to current workers (augmentation, productivity gains) create political support for continued adoption.

So we face a genuinely unprecedented situation: a technology that is good for today's economy, unstoppable in its adoption, and potentially corrosive to the mechanism (education) that has historically been society's way of adapting to technological change. The right response is not panic or prohibition. It is to ask, very carefully, what the damage actually looks like.

The Real Risk: Not Replacing Teachers, But Replacing Struggle

To be clear: the risk is not that AI replaces educators. A teacher could still be in the room, doing everything perfectly (designing meaningful tasks, scaffolding inquiry, modeling intellectual character), and it might not matter, because the student never has to struggle through the higher-order thinking themselves.

The purpose of education, at its core, is to develop a person's capacity for independent judgment. Every difficult assignment, every confusing problem set, every essay that requires organizing messy thoughts into a coherent argument; these are not obstacles to learning. They are the learning. What Bjork, Bjork et al. (2011) call "desirable difficulties": tasks that feel harder in the moment but produce more durable learning, by making the brain work harder to encode what it is doing. The cognitive struggle is not a bug in the educational process; it is the process. Previous technologies could not short-circuit this. A calculator does not make your decisions. Google does not evaluate your argument. Spellcheck does not structure your thinking. But an LLM will, and it will do it convincingly enough that the student may not realize they have outsourced the very capacity education was supposed to build in them.

The performance gains are real, but they are hollow. Yan et al. (2025) apply the performance-vs-learning distinction to generative AI, and the results are alarming. Across a meta-analysis of sixty-nine studies, ChatGPT improves student "academic performance" with an effect size of $g = 0.7$—a large effect (Deng et al. 2025). But these gains likely reflect immediate task success, not durable learning, because the two concepts are consistently conflated in the literature. In a peer feedback study, students showed significant quality improvements with AI assistance that collapsed when the AI was removed (Darvishi et al. 2024). Students who relied heavily on ChatGPT for research exhibited weaker argumentation and reasoning skills than those using traditional, cognitively demanding methods (Stadler, Bannert, and Sailer 2024). Frequent AI chatbot users reported lower learning autonomy (Zhai, Wibowo, and Li 2024). The pattern is consistent: AI tends to make the work look better without reliably making the learner stronger.

Fan et al. (2025) designed the experiment that makes this mechanism precise. In a randomized controlled trial, 117 university students were assigned to four conditions: ChatGPT support, human expert support, a checklist tool, and no support. The AI group's essay scores improved significantly ($p = 0.037$, $\eta^2 = 0.108$). But knowledge gain showed no significant difference across groups. Neither did knowledge transfer. Neither did intrinsic motivation. Process mining of learning traces revealed why: AI learners' self-regulated learning was tightly coupled to a ChatGPT consultation loop, with reduced reading, less self-evaluation, and minimal independent planning. Students performed better but learned nothing more. Fan et al. name this phenomenon "metacognitive laziness": learners quietly outsource the self-monitoring and self-evaluation that normally drive deep learning, which leaves the task complete but the learner unchanged.

This is one mechanism, and a well-documented one. AI reduces the cognitive load that feels burdensome, but that burden is precisely what drives durable learning. Students perform better with AI and worse without it. Augmentation of output does not equal augmentation of learning. The teacher can still be in the room, the assignments can still be meaningful, and the damage still compounds, because learning has been quietly outsourced. What remains is to ask, for any given student using AI, whether any learning is actually happening at all.

Part III: The Reverse Alignment Problem

AI alignment asks how to make AI systems reflect human values. This Part examines the mirror problem: how to align education with a world in which AI has absorbed much of the cognitive work education was built to develop. For the first time in this pattern, much of the working population is facing cognitive systems capable of performing many of the tasks education was built to produce, and the question is no longer only whether the machines behave well toward us, but whether the educational process that forms us still produces the capacities we need.

Is Any Learning Actually Happening?

When a student uses AI, three fundamentally different things might be happening. Option 1: the student is learning the domain skill directly, using AI as a reference (like a textbook that answers questions). The AI is a tool for content delivery. Option 2: the student is learning to delegate to AI: prompt engineering, task decomposition, and output evaluation. This is a genuine skill, and it may be valuable in the workplace. But it is a fundamentally different skill from domain expertise. A student who excels at directing AI to solve physics problems may understand very little physics. Option 3: the student is developing understanding through the act of articulating to AI: deepening understanding by being forced to explain, specify, question, and refine. The dialogue with AI becomes a form of productive intellectual struggle.

These three options are not mutually exclusive, but they have very different implications for education. If Option 1 dominates, AI is a fancier textbook. If Option 2 dominates, education should pivot to teaching AI delegation. If Option 3 is real, then AI may be a powerful learning tool, but only under conditions we need to understand.

The evidence for Option 1 is mixed. The Student Report shows nearly half ($\sim$47%) of interactions are Direct (students seeking answers, not understanding). Lodge et al. (2024) argue that while "content knowledge is clearly necessary for critical thinking," it is "not sufficient." GenAI "draws exclusively on existing data" and cannot generate genuinely new knowledge. Students using AI purely for content delivery will not develop original domain mastery. The evidence for Option 2 is clearer: Massenkoff et al. (2026) shows high-tenure users achieve 6.4 percentage points higher success rates. People learn to use AI better over time. But this is skill at AI delegation; it tells us nothing about domain learning.

Option 3 is the most intriguing and the most fragile. Gonsalves (2024) identifies two distinct forms of critical thinking in AI-assisted learning: "Critical Thinking TOWARD AI" (interrogating, validating, and refining AI outputs) and "Critical Thinking FOR assignments" (applying insights to tasks). In her study, students who actively validated AI outputs against external sources showed genuine engagement consistent with Option 3. But this was a small qualitative study ($n = 8$).

Alsaiari et al. (2026) reveal this difficulty in a larger experiment. A semester-long randomized controlled trial ($n = 329$) compared three types of AI-generated feedback: directive (specific instructions), metacognitive (reflective prompts), and hybrid (both). Purely metacognitive feedback (the kind most aligned with Option 3) produced the least revision behavior. Students receiving metacognitive prompts alone were three times less likely to revise than those getting directive feedback (OR $= 2.93$, $p = .009$). Hybrid feedback produced the highest revision rates (27.5% versus 12.1%). The interpretation: metacognitive engagement through AI is real but may overwhelm novice learners. It works best when scaffolded with directive guidance, not when the student is left to reflect alone with the machine.

This creates a bootstrapping problem: to develop understanding through articulation to AI, students may need the critical thinking skills that Option 3 is supposed to develop. Lodge et al. (2024) make this explicit: using AI effectively requires students to "critically think about what critical thinking the machine should do." Without pre-existing metacognitive capacity, the articulation process may produce cognitive overload rather than cognitive growth.

Liang et al. (2026) confirm this three-way structure is real, not merely theoretical. Their systematic review of fifty-six empirical studies, analyzed through a human–AI interaction framework, identifies three interaction modes: externalization (high AI automation, low human control; corresponding to Options 1 and 2), internalization (high human control, low AI automation; closest to Option 3), and hybrid intelligence (high on both dimensions). Internalization (where the student retains control and AI provides insight for reflection) showed the most robust positive learning outcomes. But internalization was also the hardest to implement and the slowest to show results. The field gravitates toward externalization because it is easier to build and measure.

One concrete design for Option 3 already exists. Mollick and Mollick (2023) propose "AI as Student," where the learner teaches the AI a concept and then evaluates the AI's explanation for accuracy and completeness. This inverts the typical dynamic: instead of asking AI for answers, the student must articulate, explain, and judge. The learning science is well-established: teaching others is among the most powerful techniques for deepening one's own understanding. Whether this transfers to teaching an AI remains an open question.

If So, Where Is the Ceiling?

If Option 3 is real, if students genuinely develop understanding through articulation to AI, the next question is: to what ceiling? The Anderson and Krathwohl (2001) knowledge dimension provides the framework. The four knowledge types suggest different ceilings.

Factual knowledge (terminology, specific details): AI can almost certainly scaffold learning here. Articulating factual questions to AI forces the student to organize their thinking. The ceiling is likely high, roughly on par with what previous technologies already achieved.

Conceptual knowledge (categories, principles, theories, models) is harder. AI can explain concepts beautifully, but genuine conceptual understanding requires integrating ideas into one's own mental framework. This is where the gap between perceived and actual understanding likely lives: students may acquire the feeling of conceptual understanding without the reality.

Procedural knowledge (methods, techniques, criteria for when to use them) is harder still. AI can demonstrate procedures and walk students through methods, but procedural mastery requires doing, and knowing when to apply which procedure. A student who watches AI solve differential equations may understand the steps without developing the ability to choose the right method for a novel problem.

Metacognitive knowledge (awareness of one's own cognition: knowing what you know, knowing which strategy to use, knowing when you are confused) is the highest knowledge type and arguably the most important for independent judgment. Metacognitive knowledge is what allows an expert to say "something feels off about this result" (what Schwartz calls "taste"). It is what enables Polanyi's "we know more than we can tell" (Polanyi 1966). Whether metacognitive awareness can develop through AI interaction is the critical frontier, and the ceiling here is likely lower than the others, because metacognitive knowledge tends to develop through struggling, failing, and learning to monitor your own thinking.

Gonsalves (2024) proposes "melioration" as a metacognitive skill for the AI era: information melioration (integrating AI data with validated sources) and tool melioration (combining AI tools with traditional methods). This is essentially a metacognitive skill that sits near the proposed ceiling; it requires students to know enough to judge when AI output needs supplementation. But melioration presupposes the metacognitive awareness to recognize when AI output is insufficient. For novices, that recognition may not yet exist. The feedback RCT (Alsaiari et al. 2026) confirms this: when students received purely metacognitive AI feedback, many reached what the authors describe as a "cognitive saturation point"; they invested processing effort but could not translate it into action.

Frey's Polanyi's Paradox, applied to learning, sharpens the point. Frey argues that tacit knowledge protects jobs from automation because "we know more than we can tell"; we cannot codify it, so computers cannot replicate it. But here is the twist for education: AI does not need to replace tacit knowledge. It just needs to prevent it from forming. If students skip the productive struggle where tacit judgment develops, they never acquire the very skill that Frey says will protect them in the labor market. The economic prescription (develop skills AI cannot replicate) depends on an educational process that the technology is quietly eroding.

And How Would We Even Measure It?

If Option 3 is happening, and it has a ceiling, the practical question becomes: how would we know? Corbin et al. (2025) call this a "wicked problem": one with no correct solution, only better or worse responses where every design decision comes with lived consequences.

Current assessment systems are designed to measure what students know, not how they came to know it. A student who delegates analysis to AI and a student who develops understanding through AI dialogue may produce identical outputs. Karamanis (2026) illustrates this vividly: two PhD students, identical papers, identical supervisor feedback, indistinguishable by every metric the academy uses to evaluate them, yet one built genuine understanding while the other remains a first-year student who shipped a product but never learned the trade. The Anthropic Educator Report shows that 48.9% of assessment is already automated. If AI is generating the work and AI is grading the work, where is the human signal of genuine understanding? AI detection has failed: Corbin et al. report that detection systems are "inconsistent, easily circumvented, and in some cases discriminatory." Containment is not the answer. The deeper problem is that assessment fundamentally translates from what students do to what teachers observe to what institutions record. AI complicates each step of this chain, making the connection between output and capability increasingly opaque. Anthropic's own observational tool, Clio (Tamkin et al. 2024), demonstrates that privacy-preserving conversation analysis at the scale of millions of interactions is technically feasible, but it too was designed to surface usage patterns rather than to measure impacts on learners' cognitive development.

Corbin et al. propose six fundamental questions that any post-AI assessment framework must address: Why do we assess? Who should be involved; whose capability does a grade describe when AI co-produces the work? What should we assess; which capacities still matter? How and when; what tasks can link student activity to defensible capability claims? Where: in what settings does assessment remain credible? And What if: how might assessment be reimagined entirely?

Several measurement approaches become concrete within this framework. Process-oriented assessment, as both Lodge et al. (2024) and Yan et al. (2025) recommend, focuses not on the final product but on the process students employ when interfacing with AI: steps of analysis, explanation, evaluation, and synthesis that are "difficult for a GenAI tool to simulate effectively." Transfer tests offer the simplest and most decisive measure: can students who learned with AI apply their understanding to novel problems without AI? The Darvishi et al. peer feedback study showed performance collapsed when AI was removed. If AI-assisted learning transfers, it is real. If it does not, students have learned to perform, not to understand. Longitudinal tracking, following cohorts across multiple courses, can determine whether AI-heavy learners in introductory courses develop equivalent expertise later, or remain dependent on AI verification.

The field's current evidence base is thin. Liang et al.'s (2026) systematic review of fifty-six empirical studies on generative AI in education found that only twenty-one percent fully met minimum effect size requirements for adequate statistical power. Yan et al. (2024) echo this in *Nature Human Behaviour*, calling for greater methodological rigor and warning of an "AI-induced performance illusion" where assistance creates false impressions of learning. The measurement problem is that the tools we are using are measuring the wrong thing.

Part IV: The Proposal

The Diagnostic Agenda

The most striking feature of the current evidence landscape is what is missing. The largest-scale studies of AI in education (including Anthropic's Student Report with 574,740 conversations, Educator Report with 74,000 conversations, and AI Fluency Index with 9,830 conversations) measure usage patterns. What tasks students delegate. How educators interact. Which behaviors correlate with iteration. None of them measure learning outcomes. The Student Report says so explicitly: "We only study what tasks students delegate to AI, not how they ultimately use AI outputs in their academic work or whether these conversations effectively support learning outcomes" (Handa et al. 2025a, p. 11). These reports represent unprecedented visibility into AI's educational footprint. But visibility into behavior is not the same as understanding of impact.

The critical gaps are systematic. No large-scale study has yet distinguished whether AI-assisted students are performing better or learning more; the Fan et al. RCT ($n = 117$) found the answer is the former, but whether this scales remains untested. The knowledge dimension of the 2D Bloom's matrix (Factual through Metacognitive) is entirely unmeasured in student AI usage data. The AI Fluency Index (Swanson et al. 2026) finds 85.7% of conversations involve iteration, but does not distinguish iteration for output polish from iteration for deeper understanding. The economic papers and education papers use overlapping methodologies but study different populations and never connect; no one has built the stock-flow bridge. Every study is a snapshot: eighteen days, seven days, a single semester. And no one is auditing what Dewey (1938) called "collateral learning": the incidental expertise that develops as a byproduct of "tedious" tasks like reviewing medical charts, conducting legal research, or debugging code.

These gaps are answerable. The conversation data exists, the frameworks exist, and the infrastructure for privacy-preserving analysis at the scale of millions of interactions has already been demonstrated (Tamkin et al. 2024). What is missing is not methodology; it is a research program that treats learning outcomes, not usage patterns, as the primary object of study.

Beyond the Ceiling: Extending the Ladder

But diagnosis alone is not a vision. The essay's historical analysis suggests something more ambitious: the pattern of adaptation may not be broken; it may just demand something harder than education has ever had to do.

Recall the historical pattern. Every time technology automated cognitive level N, education survived by retreating to level $N + 1$. Calculators automated computation; education shifted to application and problem-solving. Search engines automated recall; education shifted to evaluation and synthesis. Translation tools automated decoding; education shifted to cultural understanding and interpretation. We argued in Part II that AI blocks this retreat because it operates at the top of Bloom's Taxonomy: Create, Evaluate, Analyze. There is no higher cognitive floor to retreat to.

But this framing may be too pessimistic. Perhaps the escape route is not blocked. Perhaps the ladder needs to be extended.

The current 2D matrix (Anderson and Krathwohl 2001) has six cognitive process levels and four knowledge types. This framework was designed for an era when the highest cognitive demand on a student was to create something using metacognitive awareness. That was the ceiling of what education needed to aim for, because no technology could reach it. Now AI can. So the question becomes: what is above Create + Metacognitive? What cognitive capacities have we never needed to name explicitly, because education never needed to go there?

Several candidates emerge from the evidence in this essay. On the cognitive process dimension, beyond Create, lies something we might call judgment under genuine uncertainty: not evaluating known options, but navigating situations where the right framework itself is unknown. Lodge et al.'s (2024) "intellectual values" (clarity, relevance, and coherence applied in novel contexts) gesture at this capacity. AI can generate and evaluate options. It cannot yet determine which problem is worth solving. Schwartz's (2026) finding is instructive: AI operates at a "G2" level (competent execution of well-scoped problems), but lacks the "taste" to identify which problems matter. That taste sits above Create.

On the knowledge dimension, beyond Metacognitive, educational psychologists have already identified a higher level: epistemic cognition, which concerns reflecting on the limits, nature, and justification of knowledge itself (Greene, Sandoval, and Bråten 2016). Beyond even this lies what the literature calls epistemic identity: your orientation as a thinker, which questions you consider worth asking, which sources of evidence you trust, and what you are willing to stake an argument on. It is not just how you think (metacognition) or what counts as knowledge (epistemic cognition), but who you are as a thinker. What are your intellectual commitments? What questions drive you? What are you willing to be wrong about? Lodge et al.'s (2024) "virtues" (open-mindedness, intellectual humility, persistence, and honesty) sit here. These are not just metacognitive skills; they are character traits that determine how metacognition is exercised. AI has no character. It has no stakes in being right.

And there may be entirely new dimensions: relational knowing (knowledge that exists only in the interaction between people); social-epistemic awareness (understanding how knowledge is produced, validated, and contested within communities); embodied judgment (knowledge that requires presence, context, and physical engagement with the world).

To make this concrete, imagine a doctoral student writing on a contested question in the history of medicine. Claude can generate credible interpretations of the primary sources, each with its strongest evidence and its sharpest counterargument. The AI does in minutes what once took weeks. What she brings that the machine cannot is a reason to prefer one reading over another: what she thinks history is for, which kinds of evidence she trusts, which questions feel worth asking, and what she is willing to stake an argument on even when the machine can always produce an equally plausible alternative. The content of her thesis may be co-constructed. The intellectual identity that selects among possibilities is hers alone. That is the capacity above Create on the old taxonomy. It is not about generating more or generating faster; it is about caring enough about a question to defend a particular answer against a system that can always offer another one.

The historical pattern holds: technology automates level N, education discovers and cultivates level $N + 1$. The difference this time is that $N + 1$ does not yet exist in our taxonomies. Education's task is not just to adapt; it is to discover what it is adapting toward. And that discovery must happen fast, because students are already using AI before anyone has mapped the cognitive territory above it.

The calculator analogy is instructive. When calculators arrived, education did not abandon mathematics. It redefined what "doing math" meant: not computing answers but understanding when and why to apply computations to real scenarios. The homework changed. The expectations changed. The purpose changed. What did not change was the commitment to developing the student's capacity to think mathematically. AI demands the same kind of redefinition, but for thinking itself. What does "doing analysis" mean when AI can analyze? What does "doing synthesis" mean when AI can synthesize? The answer is not "nothing"; it is something higher that we have not yet named.

There is a partial historical precedent for the scale of what a response would require. Goldin and Katz (2008) document how the American high school movement from 1910 to 1940 was a genuinely radical expansion: communities, largely by local decision, built institutions to deliver a new tier of education to a much broader population than before. What this essay gestures toward is analogous in ambition but different in mechanism. Last time, the institutional innovation was physical: new buildings, new teachers, a new organizational layer. This time, the tool creating the problem could also serve as part of the delivery mechanism for the response, and its reach is potentially global rather than local. AI systems designed to develop judgment rather than replace it may come to function as a new institutional infrastructure of this kind, which is where the next section turns.

From Research to Design

The diagnostic and constructive agendas converge at a practical question: can AI systems be designed to develop the capacities that matter most, rather than undermining them? The evidence already points toward design principles. Hybrid feedback designs (combining directive guidance with metacognitive prompts) produce more engagement and revision than either alone (Alsaiari et al. 2026). Mollick and Mollick's "AI as Student" inverts the interaction: the learner teaches, evaluates, and judges, instead of asking AI for answers. The systematic review (Liang et al. 2026) shows internalization (high human control, low AI automation) produces the best learning outcomes. AI systems could be designed to favor internalization patterns when they detect educational use. Huang et al. (2025a) find that Claude already expresses context-dependent values across hundreds of thousands of real-world conversations, with transparency and clarity among its most stable orientations; the question is whether that value-responsiveness could be calibrated for educational contexts specifically, favoring scaffolding over answer-delivery when the interaction pattern suggests a learner. Gonsalves's melioration could be prompted by the system itself: "I've given you a synthesis. Before you use it, what would you check it against?"

If the diagnostic research identifies where the ceiling is, and the constructive direction maps what is above it, then design translates both into systems that push students upward rather than letting them settle into the comfortable lower-left corner of the 2D matrix. The tool that creates the problem could also be the tool that helps solve it, but only if the design is informed by learning science, not just user engagement metrics.

These agendas are deliberately sequenced. The logical next horizon is curricular reform: restructuring what is taught so that it cultivates competencies suited to an AI-mediated economy, not competencies that AI has already absorbed. But identifying which competencies genuinely complement AI requires the learning outcome data proposed above. Measure first, then redesign. The research agenda this essay proposes would produce the evidence base that curricular decisions currently lack.

What Is at Stake

If we do not do this work, the historical precedent is clear. When the race between technology and education is lost, three generations bear the cost. Working Englishmen during the Industrial Revolution "were made worse off as technological creativity was allowed to thrive. And those who lost out did not live to see the day of the great enrichment" (Frey 2019). But this time the stakes are compounded. In Frey's account, the education system eventually caught up. Three generations suffered, but the mechanism of recovery remained intact. If AI disrupts the mechanism itself (if it produces a generation less capable of the productive struggle through which expertise forms), then the recovery may not come. Not because the technology is malicious, but because we never asked, while we still could, how students were actually learning.

The fifty-five percent of professionals who express anxiety about AI (Handa et al. 2025b) are sensing something that the economic data does not capture. The Anthropic engineers who report atrophy of deeper skillsets from reduced hands-on problem-solving, and the loss of incidental learning that happens during manual work (Huang et al. 2025b), are experiencing the education problem from the inside. If the makers of AI worry about their own skill atrophy, the concern is load-bearing.

But there is also reason for hope, if we act with intent. The historical pattern of adaptation has survived every previous technological revolution. Each time, education discovered cognitive capacities it had not known it needed to cultivate. The printing press revealed that interpretation matters more than memorization. Calculators revealed that problem-framing matters more than computation. Search engines revealed that source evaluation matters more than information retrieval. AI may reveal something deeper still: that domain knowledge, the traditional product of education, was never the scarce resource. What is scarce is judgment, character, epistemic identity, and the motivation to direct knowledge toward something worth doing. These capacities have always been there, latent and unnamed, waiting for a technological pressure that forces us to cultivate them deliberately.

Centuries of educational progress have been about expanding what nurture can do: developing cognitive abilities that are not innate, through practice, mentorship, and productive struggle. The risk is not that nurture becomes futile, but that we aim it at the wrong target. When AI narrows the knowledge gap between a person and a field, what remains is the question of who that person is: what they care about, how they exercise judgment, whether they can direct their own learning. A person with natural patience and curiosity for teaching, but without formal credentials, has always been blocked by the knowledge barrier. A person with deep technical drive, but from an underrepresented background with no access to the right training pipeline, has always been blocked by the expertise barrier. AI can lower these barriers. But it can only do so for people who have developed the deeper capacities: the motivation to pursue something, the judgment to evaluate what AI gives them, the character to persist when the work is hard. If AI also short-circuits those capacities, by replacing the productive struggle through which they form, then people's functional ceiling shifts closer to whatever innate capacities they started with. The perennial "nature versus nurture" question is not settled by AI. It is reframed: nurture must shift from building knowledge to building the capacities that make knowledge meaningful.

That shift is not inevitable. But making it requires two things: the humility to measure what we do not yet understand, and the ambition to imagine what education becomes when its purpose extends beyond the knowledge that AI can now provide. The ladder has always been there. We just have not needed to climb this high before.

References

Alsaiari, O.; Baghaei, N.; Lodge, J. M.; Noroozi, O.; Gašević, D.; Boden, M.; and Khosravi, H. 2026. Directive, metacognitive, or a blend of both? A comparison of AI-generated feedback types on student engagement, confidence, and outcomes. *Computers and Education: Artificial Intelligence*, 100553.

Anderson, L. W.; and Krathwohl, D. R. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition.* Addison Wesley Longman, Inc.

Appel, R.; Massenkoff, M.; McCrory, P.; McCain, M.;

Heller, R.; Neylon, T.; and Tamkin, A. 2026. Anthropic Economic Index Report: Economic Primitives.

Bent, D.; Handa, K.; Durmus, E.; Tamkin, A.; McCain, M.; Ritchie, S.; and Jones, J. 2025. Anthropic Education Report: How Educators Use Claude. *Anthropic. Available online: https://www. anthropic. com/news/anthropic-education-report-how-educators-use-claude (accessed on 15 December 2025)*.

Bjork, E. L.; Bjork, R. A.; et al. 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, 2(59-68): 56–64.

Corbin, T.; Bearman, M.; Boud, D.; Crawford, N.; Dawson, P.; Fawns, T.; Henderson, M.; Lodge, J.; Luo, J.; Matthews, K.; Nicola-Richmond, K.; Nieminen, J. H.; Pepperell, N.; Swiecki, Z.; Tai, J.; and Walton, J. 2025. Assessment after artificial intelligence: The research we should be doing. *Journal of University Teaching and Learning Practice*, 22(7).

Darvishi, A.; Khosravi, H.; Sadiq, S.; Gašević, D.; and Siemens, G. 2024. Impact of AI assistance on student agency. *Computers & Education*, 210: 104967.

Deng, R.; Jiang, M.; Yu, X.; Lu, Y.; and Liu, S. 2025. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227: 105224.

Dewey, J. 1938. *Experience and Education*. Macmillan.

Fan, Y.; Tang, L.; Le, H.; Shen, K.; Tan, S.; Zhao, Y.; Shen, Y.; Li, X.; and Gašević, D. 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2): 489–530.

Frey, C. B. 2019. *The Technology Trap: Capital, Labor, and Power in the Age of Automation*. Princeton, NJ: Princeton University Press. ISBN 9780691172798.

Goldin, C.; and Katz, L. F. 2008. *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press. ISBN 9780674028678.

Gonsalves, C. 2024. Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. *Journal of Marketing Education*, 02734753241305980.

Gordon, R. J. 2016. *The rise and fall of American growth: The US standard of living since the civil war*. Princeton university press.

Greene, J. A.; Sandoval, W. A.; and Bråten, I. 2016. *Handbook of epistemic cognition*. Routledge New York, NY.

Handa, K.; Bent, D.; Tamkin, A.; McCain, M.; Durmus, E.; Stern, M.; Schiraldi, M.; Huang, S.; Ritchie, S.; Syverud, S.; et al. 2025a. Anthropic Education Report: How University Students Use Claude. *Anthropic*.

Handa, K.; Stern, M.; Huang, S.; Hong, J.; Durmus, E.; McCain, M.; Yun, G.; Alt, A.; Millar, T.; Tamkin, A.; Leibrock, J.; Ritchie, S.; and Ganguli, D. 2025b. Introducing Anthropic Interviewer: What 1,250 Professionals Told Us About Working with AI.

Handa, K.; Tamkin, A.; McCain, M.; Huang, S.; Durmus, E.; Heck, S.; Mueller, J.; Hong, J.; Ritchie, S.; Belonax, T.; et al. 2025c. Which economic tasks are performed with ai? evidence from millions of claude conversations. *arXiv preprint arXiv:2503.04761*.

Huang, S.; Durmus, E.; McCain, M.; Handa, K.; Tamkin, A.; Hong, J.; Stern, M.; Somani, A.; Zhang, X.; and Ganguli, D. 2025a. Values in the Wild: Discovering and Analyzing Values in Real-World Language Model Interactions. In *Proceedings of the Conference on Language Modeling (COLM)*. ArXiv:2504.15236.

Huang, S.; Seethor, B.; Durmus, E.; Handa, K.; McCain, M.; Stern, M.; and Ganguli, D. 2025b. How AI Is Transforming Work at Anthropic.

Jordan, K. 2014. Initial trends in enrolment and completion of massive open online courses. *International Review of Research in Open and Distributed Learning*, 15(1): 133–160.

Karamanis, M. 2026. The machines are fine. I'm worried about us. Ergosphere Blog. Accessed: 2026-04-10.

Liang, Z.; Yang, K.; Sha, L.; Gašević, D.; Yan, L.; and Chen, G. 2026. A systematic review of generative AI in education: Empirical insights from a human–AI interaction perspective. *British Journal of Educational Technology*.

Lodge, J. M.; Ellerton, P.; Zaphir, L.; and Brown, D. 2024. Assessing in the age of AI: is critical thinking the answer? In *Artificial Intelligence Applications in K-12*, 24–37. Routledge.

Massenkoff, M.; Lyubich, E.; McCrory, P.; Appel, R.; and Heller, R. 2026. Anthropic Economic Index report: Learning curves.

Mollick, E.; and Mollick, L. 2023. Assigning AI: Seven approaches for students, with prompts. *arXiv preprint arXiv:2306.10052*.

National Council of Teachers of Mathematics. 1989. *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics. ISBN 0-87353-273-2.

Parasuraman, R.; and Manzey, D. H. 2010. Complacency and bias in human use of automation: An attentional integration. *Human factors*, 52(3): 381–410.

Polanyi, M. 1966. *The Tacit Dimension*. Garden City, NY: Doubleday.

Schwartz, M. 2026. Vibe physics: The AI grad student. Anthropic Science Blog.

Schwarz, N. 2004. Metacognitive experiences in consumer judgment and decision making. *Journal of consumer psychology*, 14(4): 332–348.

Skinner, B. F. 1958. Teaching machines. *Science*, 128(3330): 969–977.

Soderstrom, N. C.; and Bjork, R. A. 2015. Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2): 176–199.

Stadler, M.; Bannert, M.; and Sailer, M. 2024. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160: 108386.

Swanson, K.; Bent, D.; Ludwig, Z.; Dakan, R.; and Feller, J. 2026. Anthropic Education Report: The AI Fluency Index.

Tamkin, A.; McCain, M.; Handa, K.; Durmus, E.; Lovitt, L.; Rathi, A.; Huang, S.; Mountfield, A.; Hong, J.; Ritchie, S.; Stern, M.; Clarke, B.; Goldberg, L.; Sumers, T. R.; Mueller, J.; McEachen, W.; Mitchell, W.; Carter, S.; Clark, J.; Kaplan, J.; and Ganguli, D. 2024. Clio: Privacy-Preserving Insights into Real-World AI Use. Technical report, Anthropic. ArXiv:2412.13678.

Tinbergen, J. 1975. *Income Distribution: Analysis and Policies*. Amsterdam: North-Holland Publishing Company. ISBN 0-7204-3094-1.

Yan, L.; Greiff, S.; Lodge, J. M.; and Gašević, D. 2025. Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology*, 4(7): 435–436.

Yan, L.; Greiff, S.; Teuber, Z.; and Gašević, D. 2024. Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10): 1839–1850.

Zhai, C.; Wibowo, S.; and Li, L. D. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1): 28.