

AI and the Future of Education: Systemic Redesign Based on Findings from Cognitive Science and the Critical Role of Liberal Arts

Paul Tsuchido Shew

1. Introduction: From Tools to Systems

Over the past century, new technologies including film, radio, television, computers, online platforms, and MOOCs, have often promised to transform education. Yet the strongest findings show that technology or medium by itself rarely produces lasting gains in learning. Much of the apparent effect of a new medium comes from the teaching method it carries, not the technology itself (Clark 1983). In addition, schools and universities tend to adapt innovations to fit existing institutional structures and pedagogical methods, which limits their impact (Cuban 2001; Reich 2020). This article takes those lessons seriously. Together, these cautions suggest a thesis for the present article: Generative AI will improve education only insofar as institutions redesign learning systemically in light of cognitive science and humanistic ends, rather than treating AI as a convenient technological tool. (Clark 1983; Cuban 2001; Reich 2020).

Research on effective teaching points in the same direction. The practices that most reliably help students such as clear goals, timely feedback, explicit guidance, opportunities to practice, and strong teacher–student relationships, are not technological; they are features of instruction (Hattie 2008). AI can support these practices by handling routine tasks and offering individualized learning aids. But it can also make learning shallower if students rely on fluent answers without doing the mental work that creates durable understanding. Good design therefore matters. In practical terms, we should use AI to reduce unne-

essary mental effort while protecting the productive challenge that strengthens memory and understanding. The burden on educators and instructors is to use AI to reduce extraneous cognitive load while preserving the “desirable difficulties” and retrieval practice that consolidate long-term memory (Kirschner, Sweller, and Clark 2006; Roediger and Karpicke 2006).

Major policy bodies now recommend a human-centered approach to AI in education. UNESCO urges schools and universities to safeguard human agency, equity, and privacy; to validate the pedagogical fit of AI tools; and to train educators for responsible use (UNESCO 2023). The OECD likewise emphasizes that any hoped-for benefits depend on governance, educator capability, and appropriate assessment models (OECD 2023). These recommendations align with the view taken here: AI is a means to support learning and human development, not an end in itself (UNESCO 2023; OECD 2023; Selwyn 2021).

My own professional experience illustrates both the promise and the limits of educational technology. Beginning in information-technology roles at Fujitec, Earlham College, and Harvard Divinity School, I experienced successive waves of digital change: the rise of personal computers, the rapid growth of the internet, and early e-learning platforms. As an early programmer on the Moodle LMS project and later as a Moodle system administrator, and as teaching faculty with large lecture courses every semester, I saw how IT tools can open new possibilities, and how quickly those possibilities are constrained by existing pedagogical and institutional structures. These experiences led me to focus on the human side of integration, not only the technical side. Since joining a research institute in 2018 at AGU that studies AI and social change, I have worked to integrate practice with theory. As a technologist, researcher and educator, my experience leads me to believe that meaningful adoption must be

guided by cognitive science and anchored in the wider purposes of education, including the liberal arts' commitment to the holistic development of students.

This article advances two claims. First, a design-first stance is needed: align AI use with basic learning principles so that AI scaffolds thinking rather than replaces it (Kirschner, Sweller, and Clark 2006; Roediger and Karpicke 2006). Second, a system-level stance is necessary: update assessment so it continues to measure student thinking in an AI-rich environment, and re-center curricula on judgment, ethical discernment, creativity, and integrative reasoning, the core aims of a liberal education (Hattie 2008; Selwyn 2021; UNESCO 2023; OECD 2023).

II. Considerations from Cognitive Principles for AI-Era Learning

Lessons from cognitive science provide valuable insight into how humans develop intellectually and acquire knowledge and understanding. Of the many helpful insights, the following five principles seem especially helpful for education in an age of AI.

1. *Deliberation over Automation*

Large language models (LLMs) are unusually fluent: they produce coherent, well-formed prose on demand. Fluency, however, is not evidence of understanding. From a cognitive perspective, this raises a predictable pitfall. Dual-process theory distinguishes fast, heuristic System 1 processing from slow, analytic System 2 reasoning (Kahneman 2011). Fluency invites System 1 acceptance; durable learning, by contrast, depends on System 2 habits—generating reasons, checking evidence, and articulating criteria. Human-factors research names the related risk *automation bias*: when decision aids are present, people are more likely to omit necessary checks (omission errors) or to

follow an aid's recommendation despite contradictory cues (commission errors) (Parasuraman and Riley 1997; Skitka, Mosier, and Burdick 2000). Levels-of-automation models further show that as systems migrate from "information acquisition" to "decision selection," human monitoring becomes more passive unless accountability is designed into the workflow (Parasuraman, Sheridan, and Wickens 2000).

Translating these findings into course design, AI must be positioned as a prompt for deliberation, not a substitute for it. In practice, this means (a) produce-then-compare sequences: students first outline, solve, or draft without AI; only then do they compare against an AI version; (b) compare-and-defend prompts: students must specify where they accept or reject AI suggestions and why; (c) evidence tagging: claims derived from AI require citations or verification; and (d) adversarial review: peers or the instructor play the role of a skeptical reader probing AI-assisted claims. Such requirements shift students from passive acceptance (System 1) to active justification (System 2), mitigating omission and commission errors. They also cultivate calibrated trust. The learner learns when to lean on AI and when to doubt it. These design moves reflect the broader principle to preserve the learner's cognitive work at the point of judgment even when AI helps with drafting or search.

Finally, assignments should normalize disagreement with the model. LLMs are prone to confident error; requiring students to locate likely failure modes such as hallucinated sources, shallow analogies, or category mistakes, and to propose improved prompts or counter-examples converts AI from an answer engine to a thinking partner. In short, deliberation is not merely a virtue; it is a structural safeguard against automation bias and an engine of learning (Kahneman 2011; Parasuraman and Riley 1997; Skitka, Mosier, and Burdick 2000;

Parasuraman, Sheridan, and Wickens 2000).

2. Cognitive Load

Cognitive Load Theory (CLT) distinguishes *intrinsic load* (complexity inherent to the material), *extraneous load* (inefficiency introduced by presentation), and *germane load* (the productive effort invested in building schemas) (Sweller 1988; Sweller, Ayres, and Kalyuga 2011; Paas and Sweller 2012). Unguided discovery for novices typically overwhelms working memory; while structured guidance, worked examples, and gradually faded scaffolds support schema acquisition (Kirschner, Sweller, and Clark 2006). AI complicates as well as clarifies this picture. On the one hand, LLMs can reduce extraneous load by handling formatting, routine language polishing, and administrative friction. On the other hand, if AI supplies complete solutions, it can displace germane load, bypassing the very problem-solving and explanation that consolidate knowledge.

A CLT-aligned approach to AI proceeds in three moves. First, *scaffold intrinsic load*: use AI to break complex tasks into steps, but require learner production at each step. For example, prompting “Explain why this step follows,” or “State the rule used here.” Second, *minimize extraneous load* without removing essential difficulty: allow AI to generate structure (e.g., a table, outline, or template), but keep the effortful, creative thinking squarely with the student. Third, *protect germane load* by building in self-explanation and faded guidance: early phases may include AI-generated worked examples; later phases require students to complete partially worked problems and to justify each transformation. This sequencing leverages the *worked-example effect* and prevents the *expertise-reversal effect*, the tendency for heavy guidance to hinder more advanced learners, by fading AI supports as competence grows (Kalyuga et al. 2003;

Sweller, Ayres, and Kalyuga 2011).

Design pitfalls follow directly from CLT. Verbose, unfiltered AI feedback can swamp working memory; teachers should constrain AI comments to few, high-leverage points and require student response to each point. Multimodal outputs risk split-attention and redundancy if text and diagram repeat rather than complement one another; prompts should ask AI to produce complementary representations and require students to translate across them (e.g., “Turn the diagram into equations and justify each term”) (Sweller 1988; Sweller, Ayres, and Kalyuga 2011). Properly tuned, AI can become a cognitive load manager, reducing the avoidable, scaffolding the necessary, and preserving the productive (Kirschner, Sweller, and Clark 2006; Paas and Sweller 2012).

3. Retrieval and Spacing

Robust literature shows that retrieval practice, actively recalling information, produces larger, more durable gains than passive review, and that spaced practice, distributing study over time, outperforms massed sessions (Roediger and Karpicke 2006; Cepeda et al. 2006, 2008). Retrieval has advantages even over sophisticated elaborative techniques like concept mapping (Karpicke and Blunt 2011). The common thread is desirable difficulties: tasks that are effortful enough to challenge memory, yet not so hard as to discourage practice, yield stronger long-term retention (Bjork and Bjork 2011).

AI can either undermine or strengthen these dynamics. If students use LLMs to supply answers on demand, they risk fluency illusions: the smoothness of reading an AI solution masquerades as mastery. Course designs must therefore stage retrieval before reveal. A simple rule operationalizes this: Ask first, show later. Students attempt a short recall (definitions, steps, justifications) and only

then consult AI for comparison and extension to check and extend their knowledge. AI is already regularly used in educational software to operationalize spaced retrieval by generating adaptive quiz items and scheduled reminders keyed to prior errors and forgetting intervals (e.g., increasingly longer gaps after successful recalls).

Three practices anchor the section's recommendations. First, low-stakes quizzing throughout a course, with AI-assisted item generation that instructors vet and calibrate for accuracy and difficulty. Second, cumulative prompts ("Explain today's idea using last week's vocabulary") that make spacing explicit. Third, self-explanation of retrieval failures ("Why was this answer attractive? What rule did it violate?") so that feedback is not merely corrective but diagnostic. Each practice treats AI as an amplifier of retrieval opportunity, not as a shortcut around retrieval itself. The design logic aligns with this article's broader thesis: offload routine mechanics to AI and reserve the intellectual heavy lifting for the student (Roediger and Karpicke 2006; Karpicke and Blunt 2011; Cepeda et al. 2006, 2008; Bjork and Bjork 2011).

4. Metacognitive Learning

Learning improves when students do more than read or watch; they must generate (summarize, draw, explain, analogize) and monitor their own understanding. Generative Learning Theory synthesizes a family of strategies including summarizing, mapping, teaching-back, that deepen comprehension by requiring the learner to select, organize, and integrate information (Fiorella and Mayer 2015; Mayer 2021). Chi's Active–Constructive–Interactive framework predicts that constructive (e.g., self-explanation) and interactive (e.g., dialogic critique) activities outpace merely active ones such as highlighting (Chi 2009). Classical metacognition work adds the monitoring dimension:

learners need to plan, observe, and regulate their study and problem-solving (Flavell 1979; Zimmerman 2002).

AI can serve these aims if we adopt a produce–compare–explain sequence. First, students produce a response without AI (an outline, derivation, or design). Second, they compare their work with an AI response, guided by rubrics that target explanations, evidence, and structure (e.g., “Does the AI include warrants for claims? Do definitions align with course vocabulary?”). Third, they explain their accept/reject decisions (“Which two AI suggestions did you adopt, and why? Which one did you reject, and on what grounds?”). This cycle fuses generativity with metacognitive monitoring and makes reflection accountable to the task at hand.

A second role for AI is Socratic prompting. Rather than accept general “improve this” commands, instructors can require students to elicit probe questions from the model (“Ask me three questions that would reveal whether my proof misuses definition X”), then answer those probes with references to notes or texts. When students must justify changes and cite sources, AI’s suggestions become raw material for self-regulated revision, not ghostwriting. Finally, learners should maintain process logs that explicitly track prompt chains, versions, and rationales for changes; these artifacts make metacognition visible and assessable.

The key is to ensure that AI augments, rather than replaces, the generative and metacognitive acts that matter. When students draft with AI and skip justification, they get fluent text without understanding. When they generate → compare → explain, they add the missing cognitive work. This approach is consistent with the foundations cited above and the paper’s earlier emphasis on

structured guidance and deliberative accountability (Fiorella and Mayer 2015; Mayer 2021; Chi 2009; Flavell 1979; Zimmerman 2002).

5. *Ethical Anchors*

Education is not merely the transfer of information; it is a relational and moral practice. Decades of work show that teacher–student relationships and supportive classroom climates predict engagement and achievement (Wentzel 2010; Hattie 2009). The rapid arrival of online-only education and generative AI introduces opportunities and risks that must be governed by these human ends. UNESCO urges institutions to safeguard human agency, equity, and privacy, to validate AI tools for pedagogical appropriateness, and to train educators for responsible use (UNESCO 2023). The OECD similarly emphasizes that impact depends on governance, capacity building, and fit-for-purpose assessment (OECD 2023). Christian liberal arts education especially emphasizes the holistic nature of education and the role of both the learner and the school within society as a whole, highlighting the essential relational and ethical aspects of education.

At the classroom level, disclosure and dialogue matter. Students should know when and how AI is used, and they should be invited to reflect on the ethics of their own use, including issues such as accuracy, attribution, and fairness. Assignments that bring into question value trade-offs such as comparing efficiency against transparency, or privacy against personalization, prepare students for the moral questions AI raises. Instructors also need equity checks: if AI tools become integral to coursework, access and accessibility must be ensured to avoid widening the “AI divide.”

At the institutional level, ethical anchoring requires bias awareness and valida-

tion. Studies have documented demographic performance disparities in commercial AI systems (e.g., computer vision) and have analyzed the structural reasons large language models can reproduce and amplify data-level harms (Buolamwini and Gebru 2018; Bender et al. 2021). While classroom uses differ from those domains, the general lesson carries over: do not assume neutrality. Vet tools for your context; prefer uses where outputs can be checked against course materials; and keep humans-in-the-loop for consequential judgments. Finally, the end must remain clear. In a liberal-arts perspective, AI should be viewed as a means to facilitate intellectual honesty, justice, and care for others. Courses should make that ordering visible by aligning AI use with the formation of responsible learners and citizens (UNESCO 2023; OECD 2023; Hattie 2009; Wentzel 2010).

III. Empirical Evidence from Recent Case Studies

The following case studies examine concrete examples of how generative AI has been integrated into educational contexts since generative AI was popularized by ChatGPT in late 2022. Each illustrates a distinct dimension of AI-supported learning, from guided tutoring and feedback generation to large-scale instructional redesign. Together, they provide empirical grounding for the cognitive principles outlined in the preceding section and demonstrate how pedagogical design mediates AI's impact on learning outcomes. These representative cases highlight both the measurable gains and the emerging challenges that accompany AI adoption in real classrooms.

Case Study: “Wharton & British International School”

Bastani and colleagues conducted a preregistered field RCT (randomized controlled trial) in a large secondary school during Fall 2023 ($\approx 1,000$ students across grades 9–11) to test how a generative AI tutor design shapes both imme-

diate performance and subsequent learning in mathematics. Students attended four 90-minute sessions that jointly accounted for about 15% of the term's math curriculum. Classes were randomly assigned to one of three conditions during practice: (i) GPT-Base (a standard GPT-4 chat interface), (ii) GPT-Tutor (a GPT-4 interface with teacher-authored solutions and prompts engineered to give stepwise hints rather than answers), or (iii) no-AI control. Immediately after practice, all students sat a no-AI, closed-note exam on the same content. During practice, performance surged for both AI conditions: +48% for GPT-Base and +127% for GPT-Tutor versus control. On the subsequent exam, however, the GPT-Base group scored 17% lower than control, suggesting that unguided AI access can suppress skill acquisition; this negative effect was largely mitigated in GPT-Tutor, which did not show a significant exam deficit relative to control. The authors' interaction analyses indicate that students using GPT-Base frequently copied solutions, whereas GPT-Tutor usage elicited help-seeking and stepwise reasoning. (Bastani et al. 2025; Bastani et al. 2024)

Independent institutional write-ups (Wharton/Knowledge at Wharton; Penn CHIBE) summarize the same design and effects, emphasizing that guardrails such as teacher-informed prompts, structured hints, and pre-validated solutions, preserved learning while still improving practice-time performance. (Knowledge at Wharton 2024)

This study offers unusually strong causal evidence (cluster RCT, large N, authentic setting) that AI's impact depends on design that preserves the learner's cognitive work. Unguided LLM access made students faster and "better" during practice yet less capable when the model was removed. This demonstrates classic automation bias at the learning level. By contrast, GPT-Tutor operationalized several learning-science principles: (a) deliberation over automation (hints >

answers), (b) cognitive load management (teachers supplied stepwise solutions to the model to avoid hallucinations while requiring explanation from students), and (c) retrieval/effort preservation (no-AI exam immediately after practice). The design implication is straightforward: when AI is used during practice, it must constrain answer-giving and structure help-seeking to keep germane load on the learner. The paper's documented prompt architecture also models a practical pattern for instructors and institutions seeking to adopt GenAI without sacrificing learning, including teacher-curated solutions, common-error libraries, and explicit "no full solutions" instructions. Finally, the setting (a full-year curriculum slice, multiple grades) makes the external validity stronger than lab studies and provides a case to which later sections of this article will return when considering assessment redesign and institutional guardrails.

Case Study: "Harvard Physics (PS2 Pal)"

Kestin, Miller, and colleagues ran a randomized, crossover RCT in Harvard's largest introductory physics course (Physical Sciences 2, Fall 2023). Students experienced two topical units (surface tension; fluid flow) across two weeks: one week via a custom AI tutor ("PS2 Pal") used at home, and the other week via in-class active learning with human instruction. Pre-/post-tests were administered for each unit; students also reported engagement, motivation, enjoyment, and growth mindset. The peer-reviewed article (Scientific Reports, 2025) reports that students learned significantly more in less time with the AI tutor, with effect sizes estimated between 0.73 and 1.3 SD after adjusting for ceiling effects; median time-on-task with the AI tutor was 49 minutes versus ~60 minutes in class. Students also reported higher engagement and motivation in the AI condition. The tutor was not a generic chatbot: it implemented research-based scaffolding, including carefully engineered prompts, sequential

tasking, and step-by-step solutions embedded to limit hallucinations while driving productive interaction. University and education-press coverage (Harvard Gazette; Hechinger Report) communicated the headline result: students learned more than twice as much in less time with the AI tutor than in the already-strong active-learning class. (one course, ≈ 200 students, short intervention window). (Kestin et al. 2025; Harvard Gazette 2024; Barshay 2024)

This trial matters because it pits a well-designed AI tutor against a best-practice human class, active learning, and still finds larger learning gains with less time. Methodologically, the crossover design reduces between-student noise and, combined with pretests and covariates (midterm, FCI), strengthens causal inference. Substantively, the tutor’s architecture embodies precisely the learning-science levers emphasized in this article: load management (sequencing and stepwise guidance), deliberation (prompting students to reason rather than copy), and on-demand, individualized feedback that classroom pacing cannot always deliver. The result is a clear demonstration that AI can amplify, not replace, pedagogy, provided it is engineered to enact the same principles we value in human instruction. At the same time, boundary conditions remain: one institution, two short topics, and a STEM context with objective checks; transfer to other domains will require adaptation (e.g., to sustain retrieval and metacognitive demands in open-ended writing or ethics). The authors themselves stress that success hinged on careful design, including curated solutions to mitigate hallucinations, echoing this paper’s thesis that systemic redesign and principled guardrails are prerequisites for durable gains. (Kestin et al. 2025; Barshay 2024)

Case Study: School-Level Redesign at Alpha School (“2-Hour Learning”)

Alpha School’s model compresses core academics into roughly two hours each

morning using adaptive software and AI-enabled monitoring, with the remainder of the day dedicated to workshops and “life skills.” Adults are styled as “guides” rather than licensed teachers; tuition ranges from approximately \$40,000–\$65,000 depending on campus (Elwood 2025; Shamlian 2025). Public descriptions emphasize individualized pacing via adaptive platforms (e.g., IXL) and proprietary tools that track engagement (e.g., keystroke and screen-activity monitoring) to cue coaching interventions (Elwood 2025). The network has expanded from Texas to additional U.S. sites; media coverage highlights rapid growth, strong philanthropic interest, and debates about generalizability and equity. Institutional materials assert that students “learn twice as much” and score in the “top 1%,” benchmarked to NWEA assessments, though these claims are not (yet) supported by peer-reviewed evaluations (Alpha School n.d.; Elwood 2025; Shamlian 2025).

As a whole-school redesign, Alpha illustrates what it looks like to separate practice time from human-centered activities in the daily timetable. Morning academics mediated by adaptive/AI tools can in principle reduce extraneous cognitive load by structuring problems and pacing, while preserving germane load if students still justify steps and explain reasoning. The afternoon workshop block explicitly reserves time for relational, creative, and ethical formation, goals often associated with liberal education. At the same time, the model raises design and governance questions: (1) the validity of reported learning gains without independent evaluation; (2) selection effects in a high-tuition, self-selected population; and (3) privacy/agency considerations when attention monitoring is integral to classroom management. For institutions exploring timetable redesigns, Alpha is best treated as a proof of a creative structural redesign, not as settled evidence of impact. It usefully focuses attention on three practical levers: (a) load-aware orchestration of AI-mediated

practice, (b) role redefinition for adults to emphasize coaching and deliberation rather than delivery, and (c) safeguards around data, equity, and the aims of education. (Elwood 2025; Shamlian 2025; Alpha School n.d.)

Case Study — AI-Assisted Grading and Personalized Feedback in Large Classes

Heinrich and colleagues ran randomized controlled trials across four undergraduate political science courses in 2023–2024 to test whether LLM-assisted grading of short-answer questions (SAQs) could match instructor grading while scaling personalized feedback in large classes. Using GPT-4 with instructor-authored prompts (rubrics + “gold” graded exemplars), the team randomly assigned SAQ responses to human vs. AI-assisted grading and compared (a) grade discrimination (alignment with students’ broader performance), (b) regrade requests, and (c) perceived helpfulness of feedback. The sample included 271 students, 26 tests, and 3,080 SAQ responses, 88% of which were graded with AI assistance. Results show close mimicry of instructor grading and feedback at scale; importantly, the study did not attempt to estimate downstream learning gains. (Heinrich et al. 2025)

This RCT addresses a practical chokepoint: how to preserve rich, formative feedback in mass courses. It demonstrates that instructors can use rubric- and exemplar-conditioned LLM prompts to deliver comments that approximate small-class feedback without exploding workload. From a learning-science standpoint, the intervention is best viewed as a productivity amplifier that makes it possible to assign and respond to cognitively richer tasks (SAQs) at scale. But the authors are explicit that causal effects on learning—retention, transfer, or quality of reasoning—remain an open question requiring different designs (e.g., follow-up tests without AI). The study supports course-level redesign in a Tier-2 sense: it provides a tested method for integrating AI into

feedback workflows while preserving instructor standards (through prompts that embed rubrics and “gold” responses). It also reinforces the caution that feedback quality $\neq$ learning by itself; absent requirements for retrieval, explanation, and metacognitive reflection, students may attend to comments superficially. The most actionable lesson to carry forward is design-procedural: constrain the model with clearly specified criteria and exemplars, and verify outputs in early iterations before scaling. (Heinrich et al. 2025)

Case Study — LLM-Supported Patient Simulation with Structured Feedback

Brügge and colleagues conducted a double-blind randomized controlled trial to test whether adding LLM-generated structured feedback to AI-simulated patient interviews improves clinical decision-making (CDM) among medical students. All participants conducted simulated history-taking with LLM “patients”; the intervention group received structured feedback (missing questions, red flags, logic gaps), while the control group did not. After four training sessions, the feedback group significantly outperformed controls on the Clinical Reasoning Indicator – History Taking Inventory (CRI-HTI), with partial $\eta^2 \approx 0.20$ (strong effect); subdomain analyses showed gains in creating context and securing information, with no significant change in “focusing questions.” The final analyzed sample was $N = 21$, with high inter-rater reliability ($ICC = 0.924$). (Brügge et al. 2024)

Unlike many classroom deployments that track usage or satisfaction, this trial provides direct evidence of cognitive gains in a domain of reasoning. Pedagogically, the design aligns with the produce $\rightarrow$ compare $\rightarrow$ explain sequence emphasized earlier: students perform, receive targeted, structured feedback, and then revise. The effect size (partial $\eta^2 \approx 0.20$) is notable given the brief exposure (four sessions) and the small N typical of simulation studies. For

broader adoption, two qualifiers matter. First, domain structure likely helps: clinical reasoning provides stable rubrics and observable omissions (e.g., missing differential-diagnosis questions), which makes feedback more actionable than in some open-ended humanities tasks. Second, scaling requires guardrails to ensure feedback is accurate, non-hallucinatory, and aligned with curricular standards. This study strengthens the claim that AI-mediated feedback can improve reasoning when it is structured, specific, and embedded in iterative practice. (Brügge et al. 2024)

III. Discussion and Critical Considerations

A Three-Tier Adoption & Redesign Framework

The integration of generative AI is not a one-off procurement decision but a multi-level institutional change process. This is a three-tier project consisting of safe adoption, course-level redesign, and program-level transformation, each guided by learning science and bounded by ethical commitments to equity, privacy, and relational integrity.

Tier 1: Safe adoption (immediate guardrails). The first responsibility of any educational institution is to ensure that AI use does not compromise student rights, data security, or educational integrity. This requires clear policies and transparency. Students must know what forms of AI assistance are permitted, under what conditions, and with what disclosure requirements. Faculty need institutional support to articulate expectations consistently across courses. At the institutional level, safe adoption also involves data governance: protecting personal information, avoiding dependency on opaque platforms, and ensuring compliance with evolving regulations.

Ethical constraints are especially salient here. Without explicit safeguards, the

adoption of AI risks reproducing or amplifying biases embedded in training data, privileging English-speaking students, or excluding those without reliable access to technology (Williamson, 2020; UNESCO, 2021; OECD, 2023). Establishing a baseline of inclusion, through equitable access initiatives, privacy protection, and fairness audits, forms the non-negotiable foundation of responsible adoption.

Tier 2: Course-level redesign (immediate pedagogical adjustments). Once safety is in place, the next immediate priority is teaching. Here, educators must actively redesign assignments and classroom norms to align with the five cognitive principles outlined above. Faculty development is also essential. Instructors need both technical literacy and pedagogical strategies to implement these adjustments effectively. Without sustained professional development and collaborative design support, even well-conceived policies will fail to translate into practice.

The case evidence underscores why Tier-2 alignment matters. In the high-school RCT, unrestricted GPT (“GPT-Base”) boosted practice-time correctness but reduced no-AI exam performance (-17%), while a guardrailed tutor (“GPT-Tutor”) that delivered stepwise hints preserved subsequent learning (Bastani et al. 2025). This is a concrete demonstration that design choices (hints vs. answers) mediate whether AI amplifies or bypasses cognition. In Harvard’s “PS2 Pal” crossover RCT, a carefully engineered AI tutor outperformed an already strong active-learning class on two physics topics, with higher learning outcomes in less time, again suggesting that scaffolded interaction drives gains (Kestin et al. 2025). Conversely, randomized trials of AI-assisted grading/feedback show promising scaling of high-quality comments and grading fidelity in large courses, but they do not on their own demonstrate downstream

learning gains, reminding us that productivity improvements must be paired with retrieval, reflection, and revision structures to convert feedback into learning (Heinrich et al. 2025).

Tier 3: Program-level redesign (structural transformation). Lasting transformation cannot occur through course-by-course improvisation; it requires collective redesign of the architecture of learning. This involves revisiting program learning outcomes to align with skills and dispositions that remain uniquely human such as critical reasoning, ethical discernment, generative creativity, and integrative knowledge. All courses and the overall curriculum and learning goals as a whole must be intentionally structured to cultivate these outcomes in dialogue with AI tools.

At this tier, institutions must also reconsider roles and resources. Instructional designers, educational technologists, and assessment specialists need to be integrated into faculty work, not as peripheral consultants but as co-creators of curriculum. Governance structures should include ongoing quality assurance cycles to evaluate the impact of AI integration, with quality-assurance cycles; and incentive structures (release time, recognition, funding) that make redesign feasible.

Across tiers, ethics, equity and human flourishing remain standing concerns. AI integration must never compromise privacy or consent, must actively counteract bias and inequity, and must strengthen rather than weaken the relational character of education. In practice, this means evaluating each new tool or practice not only for its technical utility but also for its impact on fairness, inclusion, and the human purposes of education. In short, technology is best used as an instrument to expand human presence, not a substitute for it

(UNESCO 2023; OECD 2023).

The three tiers scaffold systemic redesign. From baseline safety, to pedagogical renewal, to structural transformation, each can help systemic redesign align with cognitive principles and ethical aims. Adopting AI without this layered framework invites cycles of fragmented pilots and disappointing returns; adopting it within this architecture positions AI meaningfully advance genuine transformation of education. (Clark 1983; Cuban 2001)

Assessment That Still Measures Student Cognition

Assessment is where institutional intentions either reinforce or undermine the principles above. Predictably, students will use AI regardless of whether schools adapt or not. If assignments and tasks reward products that AI can trivially generate, students will become proficient at outsourcing learning with no benefit to themselves. If tasks require deliberation, retrieval, creativity, and relational accountability, AI can be harnessed to strengthen rather than weaken learning.

1) Prefer authentic and in-class demonstrations. Oral defenses, class discussions and debates, lab practicums, design critiques, and other activities demand adaptive explanation and interpersonal responsiveness that AI cannot easily mimic (Black and Wiliam 2009). 2) Elevate process evidence. Multi-stage submissions (drafts, annotated bibliographies, peer reviews, and reflective commentaries) document evolving thought and make it harder to outsource cognition; they also align with generative and metacognitive learning (Fiorella and Mayer 2015; Flavell 1979; Zimmerman 2002). 3) Re-weight rubrics. Reward clarity of reasoning, originality, synthesis, question quality, and reflective self-assessment; deemphasize surface polish that AI supplies easily. 4)

Conduct assessment audits. Program-level reviews identify tasks vulnerable to automation, promote equity across courses, and embed redesign into quality-assurance cycles (OECD 2023).

The case studies clarify why these moves are necessary. The Wharton field RCT shows that unguided AI use can depress no-AI exam scores; i.e., traditional take-home assignments may cease to measure student cognition if they permit unrestricted AI (Bastani et al. 2025). The Harvard physics crossover RCT shows that well-engineered tutors can increase learning quickly, but the measurement of that learning depended on no-AI pre-/post-tests, underlining the design principle: assessments must be able to stand without AI (Kestin et al. 2025). The BMC Med Educ RCT demonstrates that structured feedback embedded in iterative practice can improve reasoning, suggesting that assessment should capture growth across cycles, not only end products (Brügge et al. 2024). Finally, the PLOS ONE RCT on AI-assisted grading/feedback reminds us that scaling feedback fidelity is not the same as causing learning; to attribute gains to feedback, we must pair it with retrieval-based checks and delayed tests (Heinrich et al. 2025).

In brief, the assessment agenda is to measure cognition and intellectual growth, not tooling or output. Doing so requires authentic demonstrations, process visibility, rubric realignment, and periodic audits, practices already consistent with cognitive science and international guidance (UNESCO 2023; OECD 2023).

The Liberal Arts Imperative in an AI Society

The rise of artificial intelligence invites fundamental questions about what kind of education humans still need. If machines can generate essays, solve problems, and deliver information instantly, some might conclude that liberal arts

education, sometimes dismissed as abstract or impractical, has little place in an AI-driven world. In fact, the opposite is true. Precisely because AI now automates many technical and informational tasks, the human-centered competencies cultivated by liberal arts education become more essential. (Nussbaum 2010; Delbanco 2012)

As articulated in earlier analysis, durable learning depends not merely on information delivery but on deliberation, retrieval, reflection, and integration. These processes require distinctly human engagement. Liberal arts education is uniquely positioned to reinforce such engagement by cultivating critical reasoning, ethical discernment, empathy, creativity, and adaptability. These are not peripheral skills; they are the very qualities that enable students to remain autonomous thinkers and moral agents in a technological society.

Critical Thinking and Cognitive Autonomy: At its core, liberal arts education develops the intellectual habits of questioning assumptions, evaluating evidence, and engaging ideas across disciplines. Philosophy, literature, history, and theology explicitly train students in rigorous analytical reasoning. Such habits are indispensable in an environment saturated with AI-generated information of variable quality. Without the capacity to interrogate sources critically, students risk succumbing to cognitive dependence, outsourcing judgment to algorithms rather than cultivating their own. Liberal arts instruction provides the cognitive resilience necessary for genuine autonomy.

Creativity and Innovation: Beyond critical thinking, the arts and humanities foster creativity and imaginative problem-solving. These disciplines encourage students to explore open-ended possibilities, draw unconventional connections, and innovate across diverse contexts. AI can generate content, but it cannot

replicate the human capacity for inspiration, intuition, or aesthetic judgment. By cultivating these creative capacities, liberal arts education ensures students are not merely consumers of technological outputs but active contributors of new ideas and cultural meaning.

Ethical Discernment and Moral Agency: Liberal arts education also sustains what may be one of humanity's last irreducibly human domains: ethical discernment and moral agency. Through philosophy, theology, and ethics, students learn to navigate complex moral questions and consider the social consequences of technological adoption. AI is notoriously poor at providing wisdom or moral direction. In this regard, human agency must be cultivated, not assumed. Without intentional moral formation, the risk is not that machines will become more like humans, but that humans will become more like machines, and in the future, easily shaped by advanced AI. Classic liberal arts education preserves the educational value of dialogue, interpretation, and empathy that keep human learning humane (Taylor 2018; Nussbaum 2010). In the Christian liberal arts tradition, such as that of Aoyama Gakuin University, this means grounding education in values of justice, stewardship, compassion, and service. Students are invited not only to master tools but to discern the ends to which those tools ought to be directed.

Cross-Disciplinary Integration: Equally important is the liberal arts capacity for integration. AI will continue to accelerate the blurring of disciplinary boundaries, creating new hybrid fields and unforeseen challenges. The liberal arts tradition equips students with the intellectual agility to connect insights from science, technology, the humanities, and the social sciences. Such integrative thinking is vital for navigating complexity and uncertainty in an AI-saturated world.

Christian Liberal Arts as Human Flourishing: For a Christian liberal arts institution like Aoyama Gakuin, this imperative is not merely pragmatic but intrinsic to our institutional values. Education is understood as the formation of the whole person before God: mind, heart, and spirit. By cultivating intellectual rigor, ethical discernment, creativity, and service, liberal arts education prepares students not only for professional competence but for human flourishing both domestically and internationally. In this light, AI becomes a means rather than an end. AI is a tool that can extend human capability, but one that must always be subordinated to the higher purposes of love, justice, and stewardship.

In sum, AI increases the importance of liberal arts education. Its capacity to foster critical thinking, creativity, ethical reasoning, interdisciplinary integration, and moral responsibility ensures that AI serves human flourishing rather than undermines it. The Christian liberal arts tradition underscores this point by insisting that education is not only about knowledge but about wisdom. In the AI society now emerging, these commitments are indispensable.

IV. Conclusion: Human Flourishing as End; AI as Means

This article argues that generative AI will advance education only to the extent that institutions redesign learning systemically around what enables human understanding and formation. Technology, considered as medium alone, rarely delivers durable gains; effects arise from pedagogy and from the institutional architectures that sustain it (Clark 1983; Cuban 2001). In an AI-saturated environment, therefore, the essential task is not to decide whether to use AI but how to re-engineer curricula, assessment, and roles so that AI scaffolds deliberation, retrieval, and ethical agency rather than replacing them. In short: we must hold fast to the ultimate goal of human flourishing and treat AI as a con-

tingent instrument for serving it.

Three bodies of evidence converge on this claim. First, the cognitive sciences delineate the conditions under which minds actually learn: (i) deliberation over automation, to counteract fluency illusions and automation bias; (ii) load-aware design that reduces extraneous burden while preserving the “desirable difficulties” that consolidate schemas; (iii) retrieval and spacing to stabilize memory; and (iv) generative and metacognitive activity that makes students’ reasons and criteria explicit (Kirschner, Sweller, and Clark 2006; Sweller, Ayres, and Kalyuga 2011; Bjork and Bjork 2011; Roediger and Karpicke 2006; Fiorella and Mayer 2015). Second, recent field experiments demonstrate that design, not mere access, determines outcomes. Unguarded access to a fluent model can inflate practice performance while suppressing later no-AI competence (Bastani et al. 2025). By contrast, a research-engineered tutor that structures stepwise reasoning can outperform already-strong human active learning with less time on task (Kestin et al. 2025). AI-assisted grading can scale rubric-aligned feedback without collapsing instructor standards, though feedback alone does not guarantee learning absent retrieval and revision structures (Heinrich et al. 2025). Structured LLM feedback within iterative simulation can measurably improve domain reasoning (Brügge et al. 2024). Third, policy guidance converges on a human-centered stance that safeguards equity, privacy, transparency, and educator capability (UNESCO 2023; OECD 2023; Selwyn 2021).

Accordingly, a three-tier redesign agenda is warranted. Tier 1, Safe adoption, establishes non-negotiable guardrails: clarity about permitted uses, disclosure, and data governance; and baseline equity of access. Tier 2, Course-level redesign, aligns tasks with learning science including the “produce → compare →

explain” workflows; minimize extraneous load while preserving germane effort; require retrieval-before-reveal; and make metacognition assessable through process artifacts. Here the case evidence is decisive: guardrails that convert chat into effective tutoring (hints, curated solutions, accountability prompts) protect learning outcomes; generic answer-giving does not (Bastani et al. 2025; Kestin et al. 2025). Tier 3, Program-level transformation, re-centers outcomes on what remains distinctively human such as judgment, ethical discernment, creativity, integrative reasoning, and restructures roles and quality-assurance cycles so that the institution supports effective and appropriate learning models (Hattie 2008; OECD 2023).

Assessment is a linchpin. In a world where students will use AI regardless of policy, educators must design assessments that still measure student cognition: authentic oral and in-class tasks; visible process evidence across drafts and peer reviews; rubric weights that privilege reasoning, synthesis, and originality over surface polish; and periodic audits to identify automation-vulnerable tasks (Black and Wiliam 2009; Roediger and Karpicke 2006; OECD 2023). The empirical studies considered here repeatedly measured learning with no-AI pre/post instruments. An implied principle worth making explicit is that even when AI mediates practice, assessment must be able to stand without AI to ensure that curricula are cultivating knowledge and judgment that are resident in the learner (Bastani et al. 2025; Kestin et al. 2025).

Finally, the liberal arts, and especially in their Christian articulation, give clear purpose and direction to technical redesign. AI can speed the flow of information and automate routine work. That reality increases, rather than lessens, our responsibility to form people who can interpret, evaluate, and steer these powers toward the good. Liberal learning cultivates precisely these capacities:

disciplined inquiry across fields, creative imagination, empathy and moral agency, and the integrative wisdom that joins knowledge to service in community (Nussbaum 2010; Delbanco 2012; Taylor 2018). In a Christian liberal-arts frame, these aims are oriented by love of God and neighbor and expressed through justice, stewardship, and grace. On that basis, AI must remain a servant rather than a master: it is welcomed where it enlarges human presence and participation, and should be minimized where it diminishes agency, privacy, equity, or the relational heart of education (UNESCO 2023; OECD 2023).

In sum, AI will not redeem or ruin education by itself. It will do what our designs and our commitments permit. If we adopt a layered architecture of safe adoption, course-level redesign, and program-level transformation; if we anchor pedagogy in the cognitive mechanisms by which people actually learn; and if we keep the ends of liberal education including wisdom, justice, creativity, and empathy, clearly before us, AI can become a servant of human flourishing rather than its substitute. That is the criterion by which adoption should be judged.

References

- Alpha School. n.d. "Alpha School: AI-Powered Private School." Accessed October 8, 2025. <https://alpha.school/>.
- Barshay, Jill. 2024. "An AI Tutor Helped Harvard Students Learn More Physics in Less Time." *The Hechinger Report*, September 16, 2024. <https://hechingerreport.org/proof-points-ai-tutor-harvard-physics/>.
- Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. 2024. "Generative AI Can Harm Learning." Working paper. Accessed October 8, 2025. https://hamsabastani.github.io/education_llm.pdf.
- Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. 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>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell.

AI and the Future of Education: Systemic Redesign Based on Findings
from Cognitive Science and the Critical Role of Liberal Arts

2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. <https://doi.org/10.1145/3442188.3445922>.
- Bjork, Elizabeth L., and Robert A. Bjork. 2011. "Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning." In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, edited by Morton Ann Gernsbacher, Richard W. Pew, Louis M. Hough, and James R. Pomerantz, 56–64. New York: Worth.
- Black, Paul, and Dylan Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21 (1): 5–31. <https://doi.org/10.1007/s11092-008-9068-5>.
- Brügge, Emilia, Sarah Ricchizzi, Malin Arenbeck, Marius Niklas Keller, Lina Schur, Walter Stummer, Markus Holling, Max Hao Lu, and Dogus Darici. 2024. "Large Language Models Improve Clinical Decision Making of Medical Students through Patient Simulation and Structured Feedback: A Randomized Controlled Trial." *BMC Medical Education* 24 (1): 1391. <https://doi.org/10.1186/s12909-024-06399-7>.
- Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15. <https://proceedings.mlr.press/v81/buolamwini18a.html>.
- Cepeda, Nicholas J., Edward Vul, Doug Rohrer, John T. Wixted, and Harold Pashler. 2008. "Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention." *Psychological Science* 19 (11): 1095–1102. <https://doi.org/10.1111/j.1467-9280.2008.02209.x>.
- Cepeda, Nicholas J., Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. "Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis." *Psychological Bulletin* 132 (3): 354–80. <https://doi.org/10.1037/0033-2909.132.3.354>.
- Chi, Michelene T. H. 2009. "Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities." *Topics in Cognitive Science* 1 (1): 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>.
- Clark, Richard E. 1983. "Reconsidering Research on Learning from Media." *Review of Educational Research* 53 (4): 445–59. <https://doi.org/10.3102/00346543053004445>.
- Cuban, Larry. 2001. *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press.
- Delbanco, Andrew. 2012. *College: What It Was, Is, and Should Be*. Princeton, NJ: Princeton University Press.
- Elwood, Karina. 2025. "For \$65,000 a Year, a Teacher-less AI Private School Comes to Virginia." *The Washington Post*, August 26, 2025. <https://www.washingtonpost.com/education/2025/08/26/alpha-school-virginia-ai-education/>.
- Fiorella, Logan, and Richard E. Mayer. 2015. *Learning as a Generative Activity: Eight Learning Strategies That Promote Understanding*. Cambridge: Cambridge University

- Press. <https://doi.org/10.1017/CBO9781107707085>.
- Flavell, John H. 1979. "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry." *American Psychologist* 34 (10): 906-11. <https://doi.org/10.1037/0003-066X.34.10.906>.
- Harvard Gazette. 2024. "Professor Tailored AI Tutor to Physics Course. Engagement Doubled." September 5, 2024. <https://news.harvard.edu/gazette/story/2024/09/professor-tailored-ai-tutor-to-physics-course-engagement-doubled/>.
- Hattie, John. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge. <https://doi.org/10.4324/9780203887332>.
- Heinrich, Tobias, Spencer Baily, Kuan-Wu Chen, Jack DeOliveira, Sanghoon Park, and Navida Chun-Han Wang. 2025. "AI-Assisted Grading and Personalized Feedback in Large Political Science Classes: Results from Randomized Controlled Trials." *PLOS ONE* 20 (8): e0328041. <https://doi.org/10.1371/journal.pone.0328041>.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kalyuga, Slava, Paul Ayres, Paul Chandler, and John Sweller. 2003. "The Expertise Reversal Effect." *Educational Psychologist* 38 (1): 23-31. https://doi.org/10.1207/S15326985EP3801_4.
- Karpicke, Jeffrey D., and Janell R. Blunt. 2011. "Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping." *Science* 331 (6018): 772-75. <https://doi.org/10.1126/science.1199327>.
- Kestin, Greg, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. 2025. "AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting." *Scientific Reports* 15, no. 17458. <https://doi.org/10.1038/s41598-025-97652-6>.
- Kirschner, Paul A., John Sweller, and Richard E. Clark. 2006. "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching." *Educational Psychologist* 41 (2): 75-86. https://doi.org/10.1207/s15326985ep4102_1.
- Knowledge at Wharton. 2024. "Without Guardrails, Generative AI Can Harm Education." August 27, 2024. <https://knowledge.wharton.upenn.edu/article/without-guardrails-generative-ai-can-harm-education/>.
- Mayer, Richard E. 2021. *Multimedia Learning*. 3rd ed. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108859036>.
- Nussbaum, Martha C. 2010. *Not for Profit: Why Democracy Needs the Humanities*. Princeton, NJ: Princeton University Press.
- OECD (Organisation for Economic Co-operation and Development). 2023. *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*. Paris: OECD Publishing. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2023_c74f03de-en/full-report.html.
- Paas, Fred, and John Sweller. 2012. "An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of

AI and the Future of Education: Systemic Redesign Based on Findings
from Cognitive Science and the Critical Role of Liberal Arts

- Complex Cognitive Tasks." *Educational Psychology Review* 24 (1): 27–45. <https://doi.org/10.1007/s10648-011-9194-8>.
- Parasuraman, Raja, and Victor Riley. 1997. "Humans and Automation: Use, Misuse, Disuse, and Abuse." *Human Factors* 39 (2): 230–53. <https://doi.org/10.1518/001872097778543886>.
- Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. 2000. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30 (3): 286–97. <https://doi.org/10.1109/3468.844354>.
- Reich, Justin. 2020. *Failure to Disrupt: Why Technology Alone Can't Transform Education*. Cambridge, MA: Harvard University Press.
- Roediger, Henry L., III, and Jeffrey D. Karpicke. 2006. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17 (3): 249–55. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Selwyn, Neil. 2021. *Education and Technology: Key Issues and Debates*. 3rd ed. London: Bloomsbury.
- Shamlan, Janet. 2025. "Inside the \$40,000 a Year School Where AI Shapes Every Lesson, without Teachers." CBS News, October 2, 2025. <https://www.cbsnews.com/news/alpha-school-artificial-intelligence/>.
- Skitka, Linda J., Kathleen L. Mosier, and Mark Burdick. 2000. "Accountability and Automation Bias." *International Journal of Human-Computer Studies* 52 (4): 701–17. <http://doi.org/10.1006/ijhc.1999.0349>.
- Sweller, John, Paul Ayres, and Slava Kalyuga. 2011. *Cognitive Load Theory*. New York: Springer.
- Sweller, John. 1988. "Cognitive Load during Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–85. https://doi.org/10.1207/s15516709cog1202_4.
- Taylor, Charles. 2018. *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.
- UNESCO. 2023. *Guidance for Generative AI in Education and Research*. Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>.
- Wentzel, Kathryn R. 2010. "Social Supports from Teachers and Peers as Predictors of Academic and Social Motivation." *Contemporary Educational Psychology* 35 (3): 193–202. <https://doi.org/10.1016/j.cedpsych.2010.03.002>.

AIと教育の未来：認知科学の知見に基づく システミックな教育改革と リベラルアーツ教育の重要な役割

シュー土戸 ポール

本稿は、生成AIが教育を向上させるには、教育機関が学習科学を基礎とし、人間中心の目的を軸に据えた体系的な再設計を行う必要があると論じる。生成AIを安易に導入するのではなく、二つの連動した視点を提示する。第一に、設計を最優先する立場を取り、AIの利用を認知科学で確立された原則に従って、人間の思考を「代替」ではなく「支援」するように促すこと。第二に、システム全体の観点から評価方法や教育プログラム、教員の役割を改革し、人間固有の判断力や倫理的主体性、総合的な推論力を守ることを提案する。

論拠として、学習に関する五つの基本原則を挙げる。(1) 自動化よりも熟考を促すこと、(2) 認知的負荷を適切に管理し、「望ましい困難」を維持すること、(3) 想起と間隔反復を活用すること、(4) 生成的およびメタ認知的な学習を推進すること、(5) 倫理的で関係性に基づく学習を土台とすることである。さらに最近の五つの具体的事例を紹介する。ある高校の研究では、指導なしにAIチャットボットを使った生徒は試験で成績が低下した一方、段階的ヒント型のAI指導を受けた生徒は学習成果を維持した。また、大学の物理学入門コースの研究では、慎重に設計されたAIチューターが、質の高いアクティブ・ラーニング授業を超える成果を示し、適切なAIが良質な教育をさらに強化できることを示唆した。ほかにも、評価基準に沿ったAI採点が大規模クラスでも質の高いフィードバックを可能にする事例や、AIによる構造的フィードバックが医療分野の臨床的推論能力を改善した事例を挙げる。

こうした議論を基に、本稿は教育機関が実践すべき三段階の枠組みを提示する。第1段階（安全な導入）ではAI利用の規則を明確化し、データ保護やパ

AI and the Future of Education: Systemic Redesign Based on Findings
from Cognitive Science and the Critical Role of Liberal Arts

イアス検証を推奨する。第2段階（授業レベルの再設計）では、前述の認知原則を授業実践に反映することを提唱する。第3段階（プログラムレベルの変革）では、長期的成功のために組織全体として取り組むべき課題に焦点を当てる。

評価方法についても、学生の認知的成長を促進する具体的な提言を行う。最後に、AI時代にますます重要となる批判的思考力、創造性、倫理的推論、学際的統合力、人格形成を育む手段としてリベラルアーツ教育の意義を再確認し、適切な設計の下でAIを活用することで、人間の認知的成長と人間的な繁栄（human flourishing）を支えることが可能であることを示している。