EARLY DRAFT

The HI Manifesto for Education: Institutionalizing Societal Stewardship in the Age of AI

Authors: Jacob Sherson[a,b], Blerim Emruli[c], Rigmor Lyck Hansen[a,b]; Sanna Järvelä[d], Yoed N. Kenett[e], Frederik Brosbøl Kjeldsen[a,b] ; Camila Kølsen Petersen[b], Izabela Lebuda[f], Yaoli Mao[g], Yishay Mor[h], Andy Nguyen[d], Janet Rafner[i], Adam Vigdor Gordon; Seyedahmad Rahimi[j], Matthias Sōllner[k], Ana Alina Tudoran[a,l]; Dominik Dellemann[m] ; Steve DiPaola [n]; Florent Vinchon[o] ; Roni Reiter-Palmon[p] ; Jens Gamauf Madsen[q] ; Oday Darwich[r] ;

[a] *Center for Hybrid Intelligence,* [b] *Department of Management, Aarhus BSS, Aarhus University,* [c] *Lund University School of Economics and Management, Lund University, Sweden,* [d] *Faculty of Education and Psychology, University of Oulu ;* [e]*Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology;* [f] *Institute of Psychology, University of Wroclaw;* [g] *Autodesk, United States;* [i] *Department of Business and Management and Danish Institute of Advanced Studies, University of Southern Denmark;* [j] *Georgia Tech;* [k] *University of Kassel, Germany; Department of Economics, Aarhus BSS, Aarhus University,* [m] *vencortex®;* [n] *School of Interactive Arts & Technology, Simon Fraser University;* [o] *LAPEA, Université Paris Cité and Univ Gustave Eiffel;* [p] *Dep. of Psychology, University of Nebraska,;* [q] *IBC International Business College, Denmark,* [r] *Student Project House, ETH Zurich,*

Abstract. Generative AI creates a societal challenge: by making prediction, generation, and procedural execution abundant, it risks pushing economies toward automation-first configurations in which human involvement becomes peripheral, symbolic, or concentrated among already-expert actors. The danger is not only job displacement, but erosion of the expertise pipeline: firms can rationally reduce entry-level training and rely on AI-assisted short-term performance, while society still depends on the slow reproduction of human judgment, accountability, and tacit expertise. The Hybrid Intelligence (HI) Manifesto series proposes a proactive response: a human-centered AI transition in which machine prediction and human judgment are structurally combined, and markets, organizations, and institutions are shaped to reward human–AI complementarity rather than pure automation. This paper develops education as the capability infrastructure of that transition: the institution through which society reproduces the hybrid capabilities needed to shape AI-mediated futures. We present the **HI Educational Transformation Playbook**, a four-track architecture for moving education from reactive adaptation to societal stewardship. First, **mission and culture** are reframed: educational institutions become proactive societal actors that cultivate citizens, professionals, and institutional entrepreneurs capable of designing, evaluating, and governing human–AI collaboration. This is operationalized through the prediction–judgment frontier, a teachable procedure for identifying where AI prediction is appropriate, where human judgment is required, and where accountable ownership must remain human. Second, the **HI mindset** is developed through authorship-preserving collaboration and seamful system-shaping, operationalized by FERC (Frame–Explore–Refine–Commit), a governed interaction cycle grounded in Deweyan inquiry, Vygotskian internalization, and self-regulated learning. Third, **educational practice** is redesigned around explicit prediction points and judgment points, illustrated through bot co-design, stakeholder simulation, flawed-bot repair, and workflow re-engineering. Fourth, **learning objectives** shift from task execution toward judgment-centered augmentation: framing, validating, refining, and responsibly finalizing AI-supported work. We illustrate the playbook through two formative enactments: an institution-wide transformation of a Danish business and vocational college and a planned extension into physical making at ETH Zürich's maker space. We conclude that education is not simply where the AI transition must be managed, but one of the institutions through which society learns to steer it.

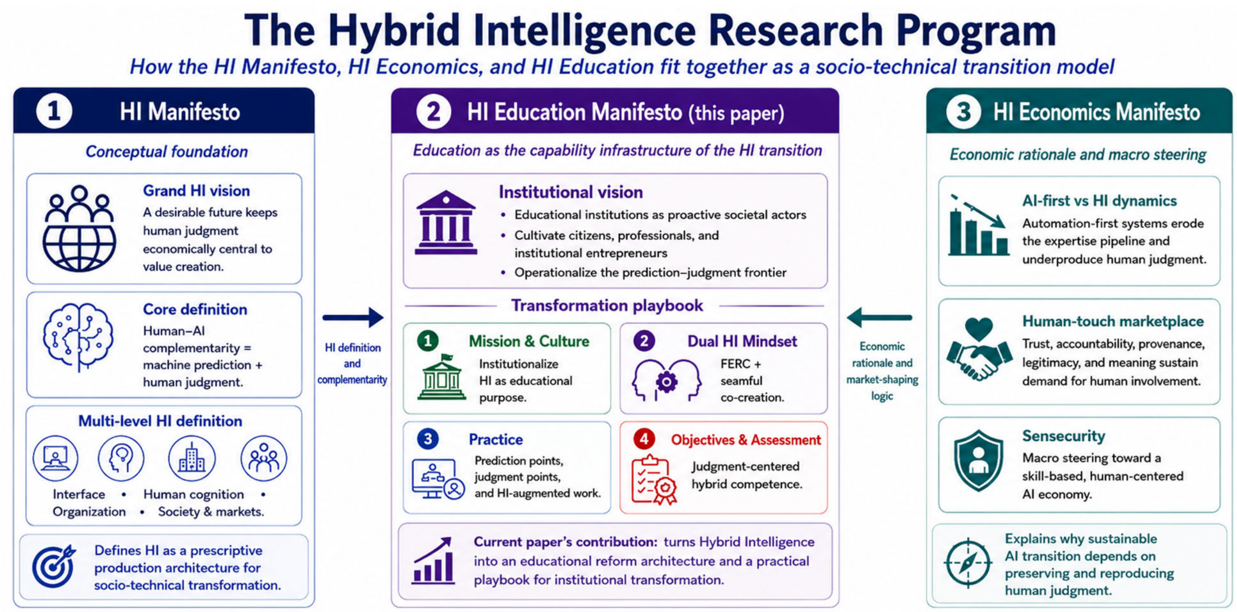
Keywords. hybrid intelligence; generative AI in education; AI literacy; human–AI collaboration; self-regulated learning; judgment; educational transformation; seamful design.

1. Introduction: Education as Societal Stewardship in the Age of AI

AI has forced education into a dilemma it cannot resolve through local policy fixes alone. In just a few years, educational institutions have moved between bans, detection regimes, honor-code revisions, AI-literacy modules, and cautious experimentation — often simultaneously, and often without a clear account of what education is ultimately supposed to become in an AI-saturated society (Holmes & Tuomi, 2022; Zawacki-Richter et al., 2019). A more constructive research agenda has emerged under the banner of Hybrid Intelligence (HI), asking not whether AI will replace or merely assist human learning, but how human and artificial cognition can be deliberately combined so that learning, judgment, and agency are strengthened rather than displaced (Cukurova, 2025; Dellermann et al., 2019; Giannakos, 2025; Järvelä, Zhao, Nguyen & Chen, 2025; Markauskaite et al., 2022). Yet the educational challenge is larger than designing better human–AI interactions. If AI is reshaping the structure of work, expertise, and value creation, education must ask a more fundamental question: how can it become the institution through which society develops the capacity to shape the AI transition rather than merely adapt to it?

This paper develops the educational pillar of a broader research program on Hybrid Intelligence as a socio-technical transition model (Sherson et al., 2026a, 2026b). The primary HI Manifesto argues that previous HI research, while foundational, has focused largely on the interaction layer of human–AI collaboration (Akata et al., 2020; Dellermann et al., 2019), and has lacked sufficiently prescriptive definitions for guiding workflow, organizational, and institutional transformation. It responds by defining HI as a production architecture in which machine prediction and human judgment are structurally combined, and by identifying education as the long-term capability pipeline without which the wider HI transition cannot be sustained (Sherson et al., 2026a). The companion economic analyses develop the demand side of this program. They argue that a sustainable AI-rich economy depends on a **human-touch marketplace**: a class of markets in which complement value — trust, provenance, accountability, liability, legitimacy, relational continuity, and meaning — sustains demand for human involvement even as AI drives the cost of functional performance toward zero (Sherson et al., 2026b). In this account, the transition to a human-centered AI economy cannot rest on sentiment or symbolic oversight alone. It requires human–AI configurations that demonstrably outperform automation-first alternatives in high-stakes domains, and a governance roadmap — termed **sensecurity** — that embeds human judgment in AI-mediated production rather than merely shielding existing jobs from automation (Sherson et al., 2026b). The present paper makes the corresponding educational argument: this societal transformation depends on whether education can continuously produce the citizens, professionals, and institutional entrepreneurs capable of building and sustaining such hybrid configurations.

The economic starting point is that AI makes prediction abundant. Pattern recognition, classification, text generation, optimization, and procedural decision-making can increasingly be executed at low marginal cost and at scale (Agrawal, Gans & Goldfarb, 2018, 2022; Eloundou et al., 2024). What remains scarce is not information, nor even competent execution, but contextual judgment: the capacity to interpret, assume responsibility, navigate uncertainty, and embed decisions within social and normative frameworks (Agrawal, Gans & Goldfarb, 2019; Acemoglu & Restrepo, 2019). The HI framework adds a crucial refinement with direct educational consequences: the boundary between prediction and judgment is not a fixed division between human and machine capacities. It is a dynamic frontier, determined by which contextual knowledge has been externalized and made operationally available to AI systems in a given socio-technical setting (Sherson et al., 2026a; cf. Polanyi, 1966). As knowledge is formalized, parts of yesterday's judgment become today's prediction. Education therefore cannot anchor its response to AI in a static catalogue of "uniquely human skills" — creativity, empathy, intuition — whose boundaries advancing systems continually erode (Rafner et al., 2021, 2023). What can be cultivated durably is the ability to operate at the moving prediction–judgment frontier: to identify, in concrete tasks and workflows, where machine prediction is useful, where human judgment remains necessary, and where accountable human ownership must be preserved.



This reframing changes education's institutional role. Historically, education has often functioned as an adaptive mechanism in technological transitions: labor markets changed, skill requirements shifted, and educational institutions supplied updated competences. That adaptive framing has always coexisted uneasily with education's formative mission — the cultivation of autonomous, responsible members of a democratic society emphasized in the Bildung-didaktik tradition (Klafki, 2000; Biesta, 2009) — and with governance regimes that have increasingly shifted from professional responsibility toward measurable accountability (Hopmann, 2008). Generative AI intensifies both tensions. It does not merely automate isolated tasks; it reshapes the architecture of problem-solving itself. Field-experimental evidence suggests that AI can significantly augment professional performance while also changing how workers structure cognition, validation, and collaboration (Dell'Acqua et al., 2023). In bounded execution tasks, AI assistance can disproportionately lift less-experienced performers, compressing skill differentials; in open-ended tasks requiring framing, domain evaluation, and creative judgment, higher-expertise workers often extract greater gains, while unstructured AI use may homogenize outputs and reduce collective diversity (Noy & Zhang, 2023; Brynjolfsson, Li & Raymond, 2025; Dell'Acqua et al., 2023; Jia, Luo, Fang & Liao, 2024; Doshi & Hauser, 2024). Read together, these findings trace the educational significance of the prediction–judgment frontier: AI devalues routine execution, but revalues the deeper, tacit, slowly acquired capacity that the expertise literature has long distinguished from procedural competence (Dreyfus & Dreyfus, 1986; Polanyi, 1966).

Labor-market dynamics compound the problem. Hiring appears to contract most visibly at the entry level of AI-exposed occupations, while experienced professionals remain comparatively resilient (Brynjolfsson, Chandar & Chen, 2025, Lodefalk, 2026). The HI economic analysis interprets this as an expertise-pipeline coordination failure: individual firms may rationally free-ride on an expertise base that the system collectively requires but that no single actor has sufficient incentive to reproduce (Sherson et al., 2026b). If AI enables acceptable short-term performance without extended apprenticeship, organizations may reduce early-career development; productivity rises immediately, while the slow formation of tacit, context-rich expertise narrows. The resulting economy risks becoming compute-based rather than skill-based: able to scale prediction and output, but unable to replenish the human judgment needed to steer, verify, contextualize, and improve AI-mediated work. Education therefore becomes the primary institutional counterweight to the short-term substitution of human learning with AI-assisted performance. It must provide not only tool fluency, but a stronger baseline capacity to learn-to-learn in AI-mediated environments, regulate one's own cognition, and participate in the redesign of hybrid workflows (Järvelä et al., 2025).

The learning stakes are already visible. Experimental studies show that unstructured generative-AI support can raise immediate performance while undermining metacognitive engagement and producing limited durable learning once support is withdrawn (Bastani et al., 2025; Fan et al., 2025). The danger is therefore double: AI can make learners look more capable while depriving them of the struggle through which capability is formed, and it can make organizations look more productive while narrowing the apprenticeship pathways through which expertise is reproduced. The educational response cannot be limited to AI literacy in the sense of knowing how AI systems work, nor to responsible tool use. Learners must develop **hybrid competence**: the ability to frame problems before prompting, interrogate and stress-test outputs, evaluate evidence and data quality, redesign workflows, and assume accountable ownership within human–AI systems. In this sense, education is not simply preparing individuals for a changing labor market. It is supplying the continual stream of hybrid-capable societal entrepreneurs on whom the transition to a skill-based, human-centered AI economy depends.

The paper's central claim follows from this diagnosis: education must move from reactive adaptation to societal stewardship. By societal stewardship, we mean the institutional capacity to shape the trajectory of technological change toward human-centered, democratically accountable, and capability-enhancing outcomes. Education is uniquely positioned for this role because it is where societies reproduce not only labor-market skills, but also judgment, autonomy, responsibility, and the ability to participate in shared world-making. The task is therefore not to defend "humanness" against machines, nor simply to add AI-literacy modules to existing curricula. It is to transform educational institutions into living laboratories for Hybrid Intelligence: settings in which learners and teachers systematically explore, design, assess, and improve human–AI collaboration before these practices harden into organizational routines across society.

This paper makes two contributions to that transformation. The first is conceptual: it reframes education as a proactive societal actor in the AI transition. Building on the HI Manifesto and its economic companion analyses, we argue that education is not downstream from technological change but constitutive of its direction. Whether the AI transition becomes automation-dominant, compliance-oriented, or genuinely hybrid depends in part on whether educational institutions can cultivate people capable of designing and governing human–AI complementarity in their local domains. This claim also reconnects economic relevance with educational formation. The capacities that a sensecurity society requires — framing, contextual judgment, accountable commitment, and the ability to redesign workflows — are not merely employability skills. They are also central to the Bildung tradition's concern with autonomy, co-determination, and responsible participation in society (Klafki, 2000; Biesta, 2009).

The second contribution is prescriptive: we develop an institution-level reform architecture for how education can assume this stewardship role. The architecture draws on the emerging HI education literature, research on self-regulated learning, human–AI collaboration, creative problem-solving, and educational transformation. At the interaction level, we use the FERC cycle — Frame, Explore, Refine, Commit — as a governed operating model that makes human–AI collaboration observable, teachable, and assessable (Sherson et al., 2026c). FERC ensures that learners articulate intent before generation, explore alternatives rather than accept single outputs, refine through comparative judgment, and commit with documented responsibility. At the design level, we introduce seamful educational design: making AI's limits, uncertainty, and handover points visible so that learners do not merely inhabit institution-designed AI environments but progressively learn to shape them. At the institutional level, we organize the transformation into four tracks: mission and culture, mindset, practice, and learning objectives. Together, these tracks move education from ad hoc AI experimentation toward a coherent institutional capacity for Hybrid Intelligence.

The remainder of the paper develops this argument in six steps. Section 2 positions the contribution against existing AI-literacy frameworks and the emerging HI education literature, arguing that current approaches provide important foundations but lack a sufficient institutional theory of education as societal stewardship. Section 3 presents FERC as a pedagogically grounded operating model for human–AI collaboration, linking it to Deweyan inquiry, Vygotskian internalization, and self-regulated learning. Section 4 formulates the grand educational challenge: what should education preserve, what should it reform, and how can institutions shift from workforce suppliers to societal change agents? Section 5 presents the HI educational transformation playbook, including seamful co-creation, institutional maturity levels, classroom activity redesign, workflow re-engineering, and the reformulation of learning objectives. Section 6 reports two formative institutional realizations — an institution-wide transformation process at IBC in Denmark and a planned extension into physical making at ETH Zürich’s Student Project House — to illustrate how the framework travels across educational levels and task domains. Section 7 concludes by outlining the research and action agenda required to evaluate, refine, and scale HI education as a mechanism for societal stewardship in the age of AI.

2. From AI Literacy to Hybrid Intelligence: The State of the Field

2.1 Two Generations of AI-Literacy Frameworks

Educational systems' first structured responses to AI arrived as literacy frameworks. A first generation — including AI4K12's Five Big Ideas (Touretzky et al., 2019), the European DigComp lineage (Cosgrove & Cachia, 2025), the UNESCO AI competency frameworks for students and teachers (Miao & Cukurova, 2024; Miao et al., 2024), and foundational academic conceptualizations of AI literacy (Long & Magerko, 2020; Ng et al., 2021) — established the field's baseline: learners should understand how AI systems work, critically evaluate their outputs, and reflect on their ethical and societal implications. These frameworks accomplished essential ground-clearing work, creating shared vocabulary and legitimizing AI as curricular content across jurisdictions.

A second generation is best represented by the joint OECD/European Union AILit Framework (OECD/European Union, 2026), which anchors the PISA 2029 Media and AI Literacy assessment and thereby carries unprecedented institutional weight. AILit marks genuine conceptual advances over its predecessors. It explicitly distinguishes AI literacy from AI tool use. Its *Manage AI* domain moves beyond understanding and evaluation toward task allocation: learners are expected to decompose problems, decide deliberately when AI should automate or augment, and establish checkpoints between human and AI roles — bringing mainstream policy closer than ever before to a hybrid-intelligence conception of competence. Its *Shape AI* domain positions learners as improvers of AI systems rather than mere consumers. And the framework is notably candid about the learning risks of unstructured AI use, acknowledging that AI-assisted performance gains often fail to translate into durable learning and citing the emerging evidence on metacognitive offloading (Fan et al., 2025; OECD/European Union, 2026). With AILit, the question is no longer whether education systems should move beyond critical consumption of AI; the flagship policy framework already says they should. The question is whether the framework's foundations can carry the ambition.

We identify two respects in which they cannot — not as failures of a primary- and secondary-education framework, but as structural limits that a hybrid-intelligence paradigm must address for education systems as a whole. The first concerns how the human contribution is defined. AILit anchors the human side of collaboration in a fixed catalogue of capacities, asserting that AI "lacks emotions, ethical reasoning, critical thinking, context and originality" and calling on learners to cultivate these distinctly human strengths. As argued in Section 1, this foundation is unstable: frontier systems progressively simulate capabilities once considered exclusively human, and empirical work already documents AI performing well in tasks previously classified as judgment-intensive while eroding others through homogenization and deskilling dynamics (Rafner et al., 2021, 2023). Curricula anchored in capability claims therefore face built-in obsolescence — each capability that migrates across the boundary invalidates the learning objectives built upon it. The hybrid-intelligence alternative developed in the companion Manifesto grounds complementarity instead in the *knowledge structure of the task*: the dynamic frontier between prediction and judgment, determined by which contextual knowledge has been externalized and made operationally available to machine systems in a given setting (Sherson et al., 2026a; Polanyi, 1966). On this foundation, the durable educational objective is not any fixed skill list but the capacity to locate and operate at a moving frontier. Crucially, this reframing is not more abstract than the capability catalogue it replaces — it is more operational. Where judging whether a task "requires creativity" or "genuine critical thinking" demands contested philosophical distinctions, the prediction–judgment explicitation reduces to a question any learner can be taught to ask of any subtask: does its solution rest on abundant, explicit, formalized data available to the system, or does it depend on implicit, tacit, contextual knowledge that has not been externalized? This transforms complementarity from an abstract definitional debate into a concrete assessment procedure that every student can apply — the first instance of a pattern this paper develops throughout: converting hybrid-intelligence principles from expert judgment calls into standard operating procedures accessible to all learners

The second limit concerns how collaboration itself is taught. AILit rightly emphasizes iteration: learners should brainstorm with AI, elicit feedback, refine outputs, and reflect on the process. Yet iteration is presented as a general principle whose quality depends on the individual learner's disposition and the individual teacher's scaffolding — a craft, unevenly distributed by design. What the framework lacks is a governed operating procedure: a specified interaction sequence, grounded in learning theory, that makes framing, comparative evaluation, and accountable commitment observable, teachable, and assessable behaviors rather than hoped-for attitudes. Decades of research on creative problem solving and design cognition specify what such a procedure must contain, and Section 3 presents one — the FERC cycle (Frame–Explore–Refine–Commit; Sherson et al., 2026c) — as this paper's operational core, turning AI collaboration from an individual craft into a standard operating procedure that all learners can follow and all institutions can assess.

Finally, the two paradigms assign learners different societal roles. AILit's culminating vision is the responsible, critical participant in a world marked by AI — a necessary civic foundation. The hybrid-intelligence framework assigns a further role: learners as apprentice workflow designers, whose AI use is deliberately oriented toward simulations of real-world practice in which the difference between automation-first and augmentation-oriented configurations becomes visible and consequential (Sherson et al., 2026a). This real-world anchoring serves a double function. Pedagogically, it grounds theoretical learning in authentic practice and supplies the motivational answer to the perennial "why are we learning this?" — the question AI-assisted shortcut strategies make acute. Structurally, it makes education a site where real-world expertise begins accumulating *before* labor-market entry — directly addressing the expertise-pipeline failure identified in Section 1. This role matures naturally in post-secondary and professional education, but, as we argue in Section 5, its elements can be integrated progressively downward through the curriculum.

2.2 Hybrid Intelligence in Educational Research

Alongside these policy developments, educational research has begun to constitute hybrid intelligence as a research agenda in its own right. Building on the foundational interaction-level HI literature (Dellermann et al., 2019; Akata et al., 2020), Cukurova (2025) articulates a vision for HI in education in which learning analytics and AI are designed to augment rather than substitute human cognition, explicitly warning against the substitution paradigm that dominates commercial AI-in-education offerings. Markauskaite et al. (2022) map the capabilities learners need for a world entwined with AI, arguing that these exceed what traditional literacy conceptions capture. Giannakos (2025) positions HI as a central lens for human learning research, and empirical work has begun to trace how hybrid collaboration actually unfolds in authentic educational tasks: Nguyen, Hong, Dang and Huang (2024) identify distinct human–AI collaboration patterns in academic writing, showing that learners differ systematically in how they distribute framing, generation, and evaluation between themselves and the system. The present paper builds directly on this agenda and extends it with the economic grounding and institutional architecture developed in the companion papers.

Within this literature, the self-regulation strand carries particular weight for our argument. Järvelä, Zhao, Nguyen and Chen (2025) conceptualize human–AI coevolution in learning through the lens of regulation: what learners must develop is not merely knowledge about AI but the metacognitive and socially shared regulatory capacity to monitor, evaluate, and control learning processes in AI-saturated environments (cf. Järvelä & Hadwin, 2013; Molenaar, 2022). The empirical warnings are now substantial. Unstructured generative-AI support induces what Fan et al. (2025) term metacognitive laziness — offloading of precisely the monitoring and evaluation processes on which durable learning depends — while producing immediate performance gains that vanish once support is withdrawn (Bastani et al., 2025) and, in broader populations, correlating with reduced critical-thinking engagement (Gerlich, 2025). The implication is one this paper places at its center: AI literacy, in the sense of knowledge and attitudes about AI, does not by itself confer regulatory capacity. Knowing that one should evaluate critically does not produce evaluation under conditions of fluent, rhetorically compelling, always-available generation. Regulatory capacity must be *built into the structure of the interaction itself* — scaffolded, practiced, and progressively internalized. Notably, AILit cites this same evidence base but offers no interaction-level mechanism to counteract the dynamics it acknowledges; supplying that mechanism, in a pedagogically grounded and institutionally scalable form, is the task of the remainder of this paper.

Taken together, the HI-education research agenda has produced a compelling vision (Cukurova, 2025), a learning-theoretical mechanism (Järvelä et al., 2025), and emerging empirical maps of collaboration practice (Nguyen et al., 2024). What it has not yet delivered — and what neither generation of literacy frameworks provides — are three layers this paper develops: a prescriptive interaction architecture that operationalizes regulation and authorship as teachable behavior (Section 3); an account of the institutional purpose and reform dilemma such an architecture confronts (Section 4); and a transformation playbook with maturity levels through which institutions can move from ad-hoc experimentation to embedded hybrid capability (Section 5).

3. FERC: A Pedagogically Grounded Operating Model for Human–AI Collaboration

The preceding section identified the gap this section fills: education needs an interaction architecture that converts hybrid collaboration from an individual craft into a standard operating procedure — one that makes framing, comparative evaluation, and accountable commitment observable, teachable, and assessable for all learners, not only the metacognitively gifted. This section presents the FERC cycle (Frame–Explore–Refine–Commit; Sherson et al., 2026c) as that architecture, and grounds it in the learning theory that an educational deployment requires.

3.1 From Prompt Craft to Governed Interaction: The FERC Cycle

Prevailing approaches to teaching AI collaboration center on prompt engineering: constructing better inputs to obtain better outputs (Giray, 2023; White et al., 2023). This framing has two educational defects. It treats the interaction as essentially one-shot, whereas consequential work with generative systems is inherently multi-round — users are routinely dissatisfied with initial outputs and struggle to improve them through subsequent turns (Kim et al., 2024). More fundamentally, it optimizes the artifact rather than governing the process, leaving unspecified who frames the problem, how evaluation reshapes subsequent search, and when and why interaction stops. When these dimensions remain implicit, authorship drifts: framing, evaluative criteria, and stopping decisions migrate gradually from learner to system, without any single moment of deliberate delegation (Sherson et al., 2026c; cf. Bozkurt, 2024).

FERC responds by specifying a governed interaction cycle with four phases. In *Frame*, the learner articulates intent before any generation occurs: goals, constraints, values, success criteria, and non-goals. In *Explore*, the learner deliberately instructs the AI to generate multiple alternative options within that frame. In *Refine*, the learner evaluates the alternatives comparatively — articulating strengths, weaknesses, trade-offs, and missing elements — and explicitly revises the criteria or instructions for the next Explore round, so that evaluation reshapes the search space rather than polishing a single artifact. In *Commit*, the learner selects a course of action, documents the rationale, and assumes accountability for the outcome. Explore and Refine iterate; Frame and Commit bound the cycle with human intent on one side and human responsibility on the other.

These four phases are not an arbitrary ordering of good habits. As developed in detail in the companion FERC paper, each phase operationalizes a governance function that decades of creative problem-solving research identify as necessary for high-quality outcomes: problem construction (Getzels & Csikszentmihalyi, 1976; Mumford et al., 1991; Reiter-Palmon & Robinson, 2009), deliberate divergence before convergence (Finke, 1996; Nijstad et al., 2010), evaluative reframing in which judgment reconstructs the search space (Kleinmintz et al., 2019; Sowden et al., 2018), and accountable closure. The cycle likewise maps directly onto the Double Diamond design process, with one FERC cycle governing problem construction and a second governing solution construction (Sherson et al., 2026c). FERC thus imports into everyday AI interaction the process structure that design education and creativity research have long prescribed for human work — now made urgent because fluent generative systems actively invite its collapse.

One requirement deserves emphasis because it illustrates the framework's operational character. Explore mandates a minimum of two alternatives — never critique of a single output. The rationale is psychological rather than stylistic: human evaluators are demonstrably better at comparative than at absolute evaluation, and a single option tends to anchor subsequent judgment rather than submit to it (Hsee, 1996). The difficulty is amplified in the generative-AI context, where the single output arrives fluent, confident, empathetically attuned, and rhetorically accomplished — a persuasive artifact optimized to be accepted. Requiring alternatives converts the learner's task from resisting a persuasive text to adjudicating between competitors, a cognitively tractable and trainable activity. Small procedural rules of this kind — teachable to any student, checkable by any teacher — are what distinguish an operating procedure from an aspiration, extending the operationalization pattern introduced in Section 2.

FERC also occupies a deliberate position between two educationally familiar extremes (Sherson et al., 2026c). Passive answer-seeking — the dominant casual usage pattern — minimizes learner effort and silently outsources framing, evaluation, and closure to the system. Full Socratic tutoring maximizes cognitive engagement but demands sustained scaffolding effort that is difficult to maintain at scale. FERC targets the region between: enough structure to keep the learner in the author's seat, light enough to be practiced in every ordinary task. For mass education, this scalability is not a convenience but a design requirement.

3.2 Grounding FERC in Learning Theory

For an educational deployment, process validity is not enough; the framework must be grounded in an explicit conception of learning. We ground FERC in three convergent traditions. Their convergence is itself the finding: a framework derived from creativity and design research turns out to recapitulate, with striking fidelity, the process structures that a century of learning theory identifies as constitutive of durable learning.

The first anchor is Dewey's account of inquiry and reflective thinking. For Dewey, learning is neither transmission nor unguided discovery but the disciplined transformation of an indeterminate situation into a resolved one: the felt difficulty is instituted as a problem, candidate solutions are generated and reasoned through, consequences are tested, and the result is a warranted judgment for which the inquirer takes responsibility (Dewey, 1933). The correspondence to FERC is close to one-to-one: Frame enacts Dewey's problem institution — the step Dewey insisted was the most educationally consequential and most often skipped; Explore generates candidate resolutions; Refine performs the reasoning and testing through which the problem and solution co-evolve; Commit issues the warranted assertion and accepts its consequences. This lineage matters practically as well as rhetorically: it positions FERC not as a novel technique imposed on education from management practice, but as a re-implementation of the inquiry cycle progressive education has pursued for a century — now enforced in contexts where a fluent generative counterpart makes skipping the problem-institution step effortless and therefore endemic.

The second anchor is the sociocultural tradition. In Vygotskian terms, AI systems are cultural tools that mediate cognition, and the design question is what forms of mediation produce development rather than dependency (Vygotsky, 1978). Two concepts carry the analysis. Scaffolding names support calibrated to the learner's zone of proximal development that is progressively withdrawn as competence grows (Wood, Bruner & Ross, 1976); the pathology of unstructured AI use is precisely scaffolding that never fades — permanent support that substitutes for, rather than develops, capability. FERC's phase structure functions as a counter-scaffold: it is the *interaction discipline itself* that is initially externalized — as explicit prompts, checklists, and interface structure — and progressively internalized until framing before generating, demanding alternatives, evaluating comparatively, and committing deliberately become the learner's own cognitive habit. This is Vygotskian internalization in its classical sense: external, socially organized regulation becoming self-regulation. The same lens grounds the design principle this paper develops in Section 5 as *seamful* educational design: rather than maximally smooth delegation, learning environments should preserve deliberate seams — designed points where machine support is withheld and learner judgment is structurally required — placing the productive struggle exactly where development happens.

The third anchor is the self-regulated learning (SRL) tradition, which supplies the mechanism connecting interaction structure to learning outcomes. SRL models describe competent learning as a cyclical activity of forethought, performance monitoring, and self-reflection (Zimmerman, 2002), extended in collaborative settings to socially shared regulation (Järvelä & Hadwin, 2013) and, most recently, to hybrid human-AI regulation in which regulatory work is distributed across learner and system (Järvelä et al., 2025; Molenaar, 2022). The FERC phases map directly onto the SRL cycle: Frame operationalizes forethought (goal setting, strategic planning); the Explore–Refine loop operationalizes performance-phase monitoring and control (strategy use, comparison, adaptive revision); Commit operationalizes self-reflection and attribution — with the addition, crucial in AI contexts, of explicit accountability. This mapping explains *why* the evidence reviewed in Section 2.2 shows what it shows: unstructured generative support removes exactly the regulatory operations SRL identifies as learning-constitutive — hence metacognitive laziness (Fan et al., 2025) and the evaporation of gains once support is withdrawn (Bastani et al., 2025) — while FERC re-inserts those operations as mandatory workflow elements. Parallel work in creative metacognition reaches the same conclusion from the creativity side: self-regulatory prompts and metacognitive structure measurably improve creative performance (Lebuda & Benedek, 2023; Zielińska et al., 2025). In SRL terms, FERC is a regulation scaffold; in Deweyan terms, an inquiry discipline; in Vygotskian terms, an internalizable cultural tool. The three descriptions pick out the same structure.

3.3 FERC as a Teachable Progression: Observable Behavior, Assessment, and Limits

What the learning-theoretical grounding buys, finally, is assessability. Because each FERC phase corresponds to observable interaction behavior — did the learner articulate criteria before generating? request and receive genuine alternatives? evaluate them comparatively and revise the brief? commit with documented rationale? — the framework supports formative and summative assessment of the collaboration process itself, independent of output quality. This matters doubly in the AI era: output-based assessment is increasingly confounded by machine contribution, whereas process-based assessment of authorship-relevant moves is not only still valid but newly central. The approach follows the logic of cognitive apprenticeship — making expert thinking visible so it can be modeled, coached, and progressively faded (Collins, Brown & Newman, 1989) — and satisfies constructive alignment, since the intended competence (governed hybrid collaboration), the learning activity (FERC-structured tasks), and the assessment (observable phase behavior) are one and the same object (Biggs, 1996). Section 5 develops the corresponding exercise formats and institutional maturity levels; interface-level support for rendering the phases visible and measurable, including the FERC-bot training environment, is presented in the companion paper (Sherson et al., 2026c).

Two boundary conditions temper the claims. First, FERC is designed for solution-oriented tasks — those culminating in a decision, artifact, or action; knowledge-acquisition and purely exploratory learning follow different interaction logics, requiring critical validation but not the full governed cycle (Sherson et al., 2026c). An educational implementation must therefore teach task discrimination — itself an application of the prediction–judgment assessment procedure from Section 2 — rather than ritualize FERC universally. Second, and more fundamentally, externalizing interaction structure does not guarantee that the corresponding cognition is exercised: a learner can traverse all four phases while tacitly adopting AI-suggested framings, criteria, and stopping points, performing authorship without enacting it. FERC should therefore be understood as a scaffold that renders authorship-relevant moves visible, inspectable, and coachable — raising the probability of genuine self-regulation without guaranteeing it. This is precisely why the framework cannot stand alone as an individual technique but requires the institutional embedding — teacher practice, assessment design, and organizational maturity — to which the remainder of this paper turns.

4. The Grand Educational Challenge: What to Reform, What to Keep?

Institutional responses to generative AI have so far oscillated between two poles: prohibition and policing on one side — a response that fails technically, as detection tools are unreliable and systematically biased against non-native writers (Weber-Wulff et al., 2023; Liang et al., 2023), and pedagogically, as it converts the teacher–learner relationship into an adversarial one — and uncritical efficiency-driven embrace on the other, which courts precisely the hollowing-out of learning documented in Section 2.2. Both poles construe AI use and learning as zero-sum, differing only in which side of the trade-off they protect; Section 3 has argued that the trade-off is not fixed but designed, dissolving the surface dilemma at the interaction level. Yet beneath it lies a harder institutional question that no interaction framework can answer by itself, and which this section confronts: all educational institutions now face the same foundational dilemma of how much of the traditional curriculum should be preserved, and how much must be fundamentally redesigned in light of generative AI. Reform too little, and education risks irrelevance — training students for workflows that no longer exist. Reform too radically or indiscriminately, and institutions may erode forms of expertise whose long-term societal value exceeds short-term market signals. The challenge is not incremental adjustment but principled calibration.

4.1 The Calibration Dilemma and the Offline-Fraction Principle

A natural starting point is professional reality. Curriculum design should be anchored in the work practices graduates are expected to enter. One pragmatic guiding question follows: what fraction of their future professional activity will actually be performed without AI support? If substantial portions of professional practice remain offline — requiring independent reasoning, embodied skill, or real-time judgment — then robust offline mastery must be retained. If, by contrast, AI-mediated workflows dominate, then curriculum must increasingly cultivate hybrid competence: the ability to frame problems, evaluate machine-generated options, refine outputs, and assume accountable ownership — the competence architecture specified in Section 3.

Yet even this seemingly practical principle raises a deeper question: which professional reality should guide us? The current job market? A forecasted future modeled by governments or consultancies? Or a future that educational institutions actively intend to help bring about? Current practice is transitional and uneven. Forecasts are historically fragile and often shaped by institutional incentives — the automation-exposure literature itself illustrates the point, with occupation-level and task-level methodologies producing estimates of at-risk employment differing by a factor of five for the same economies (see Sherson et al., 2026b). Aligning education solely with present adoption patterns or projected automation curves risks reproducing a future defined by short-term efficiency logics rather than long-term human flourishing. If institutions restrict themselves to reactive alignment, they become followers of technological momentum rather than co-authors of societal direction.

4.2 The Double Bind: Slow Organizational Diffusion, Fast Credential Depreciation

Two structural features of the current transition sharpen this dilemma into a genuine bind. The first is a robust regularity of every general-purpose technology transition: organizations are notoriously slow to transform themselves around new technological possibilities. Productive deployment requires complementary reorganization — of workflows, roles, incentives, and skills — that lags the technology itself by years to decades, as documented for electrification and computing alike (David, 1990; Bresnahan & Trajtenberg, 1995) and formalized in the productivity J-curve, in which measured gains materialize only after a prolonged period of intangible complementary investment (Brynjolfsson, Rock & Syverson, 2021). An education system that calibrates its curriculum to *current* organizational practice therefore calibrates to a lagging indicator: it prepares graduates for the pre-transformation workflows of employers who will restructure mid-career beneath them.

The second feature compounds the first. As established in Section 1, hiring is contracting precisely at the entry level of AI-exposed occupations (Brynjolfsson, Chandar & Chen, 2025): recent graduates lack the tacit, on-the-job expertise that continues to command a premium, while the academic-disciplinary competence their credential certifies is exactly what employers now perceive as most readily substituted by a capable chatbot. The combination is perverse: the labor market signals that it values what education does not yet systematically produce — accumulated judgment in AI-rich workflows — while discounting what education certifies. Educational institutions thus face a choice more stark than the calibration language of Section 4.1 suggests: continue optimizing the traditional disciplinary pipeline and risk educating into unemployment, or rethink the core mission of a learning institution altogether.

4.3 From Supplier to Change Agent: The Living-Laboratory Mission

The HI perspective therefore reframes the role of education. Educational institutions are not merely suppliers of employable labor; they are societal change agents. Their mission is to cultivate forms of expertise, judgment, and responsibility that sustain a resilient, human-centered economy under conditions of accelerating AI capability. Rather than asking only "What skills does the current market demand?", the HI approach asks: what kinds of human–AI complementarity should define the society we aim to build?

Taking this mission seriously implies a concrete institutional form: the learning institution as a *living experimentation laboratory* — a protected setting in which the consequences of emerging AI capabilities for individuals, organizations, and society are systematically explored, documented, and shaped before and while they diffuse through the slow-moving organizational landscape described in Section 4.2. On this model, the diffusion lag is converted from a threat into education's distinctive opportunity: precisely because firms reorganize slowly, institutions that experiment deliberately can accumulate and transmit hybrid-workflow expertise *ahead* of the labor market rather than behind it. The graduate this mission produces is what we term the *disciplinary chaos pilot* (Danish: *faglig kaospilot*): a professional grounded in domain expertise who is prepared not merely to survive labor-market turbulence but to transform it — carrying into organizations the capacity to redesign workflows away from AI-first automation and toward durable human–AI collaboration. This is the educational face of the role assigned to learners in Section 2.1 and enacted literally in the co-design structure of the ETH case (Section 6.2).

The manifesto series of which this paper is part exists to supply the holistic vision such a mission requires: a *technology-for-good* orientation in which "good" is defined not as ethical ornament but economically — fulfilling the HI mission of a marketplace in which human–AI collaboration becomes the effective norm and the value of human expertise appreciates through use, rather than draining away click by click (Sherson et al., 2026a, 2026b). From this standpoint, the offline-fraction principle of Section 4.1 becomes normative as well as descriptive. Curriculum should not preserve traditional content by habit, nor eliminate it on the basis of automation enthusiasm. It should be calibrated to the forms of human judgment that a sustainable hybrid society requires — whether or not these are fully priced into current labor markets. The central design question is therefore not simply how much to keep or reform, but how to deliberately structure education so that graduates can both operate within and actively shape emerging human–AI work practices.

4.4 Qualification, Formation, and the Purpose of Education

A change-agent mission articulated partly in economic terms invites a familiar and legitimate objection: that it instrumentalizes education for the labor market. The objection deserves a direct answer. Following Biesta (2009), education serves at least three functions: qualification (knowledge, skills, competence), socialization (induction into practices and traditions), and subjectification (the formation of autonomous persons capable of independent judgment and action). The AI transition destabilizes all three at once — qualification, because the competence profile the economy rewards is shifting toward judgment (Section 1); socialization, because the practices into which learners are inducted are themselves being reorganized; and subjectification, most fundamentally, because unstructured delegation to fluent systems erodes exactly the independent judgment in which autonomy consists.

It is here that the continental Bildung tradition, far from opposing the argument, completes it. In the didaktik lineage, the aim of general education (*Allgemeinbildung*) is the formation of persons capable of self-determination, co-determination, and solidarity (Klafki, 2000) — capacities exercised, not merely possessed — and such formation proceeds through engagement with the *epochal key problems* of one's time: the era-defining challenges whose understanding and governance a generation cannot delegate. Artificial intelligence is, by any measure, such a problem. On this reading, hybrid-intelligence education is not an importation of labor-market logic into the school but a contemporary enactment of the Bildung program: learners confront the defining socio-technical transformation of their epoch, develop the judgment to govern their own participation in it, and acquire the capacity to shape — rather than merely undergo — the workflows and institutions it produces. The disciplinary chaos pilot of Section 4.3 is, in this vocabulary, simply *Mündigkeit* under AI-era conditions. The convergence identified in Section 1 thus reappears at the level of purpose: the capacities the AI economy now scarcifies and rewards — problem framing, contextual judgment, accountable commitment — are the same capacities the formation tradition has always named as its end. Education does not have to choose between employability and formation; in the AI transition, for perhaps the first time in the industrial era, the two point in the same direction.

Two clarifications guard this claim against misreading. First, the convergence is contingent, not automatic: it holds only under interaction regimes that make judgment structurally necessary; AI-first educational adoption serves neither employability (Section 4.2) nor formation. Second, the framing is deliberately aspirational rather than defensive. The task is not to protect a residue of "humanness" against encroaching machines — a posture both theoretically unstable (Section 2.1) and pedagogically dispiriting — but to enable learners to become more capable, more autonomous, and more responsible *through* structured collaboration with AI than they could become without it. Formation theory itself supplies the design implication: development requires resistance. The friction that generative systems so effectively remove — the struggle of formulating, drafting, comparing, deciding — is not an inefficiency of learning but its mechanism, a point on which the didaktik tradition, the self-regulation evidence of Section 2.2, and institutional practice (Section 6.1) independently converge. The seamful design principle developed in Section 5.1 is, in these terms, the deliberate preservation of formative friction within AI-rich environments.

4.5 Accountability, Assessment, and the Governance Squeeze

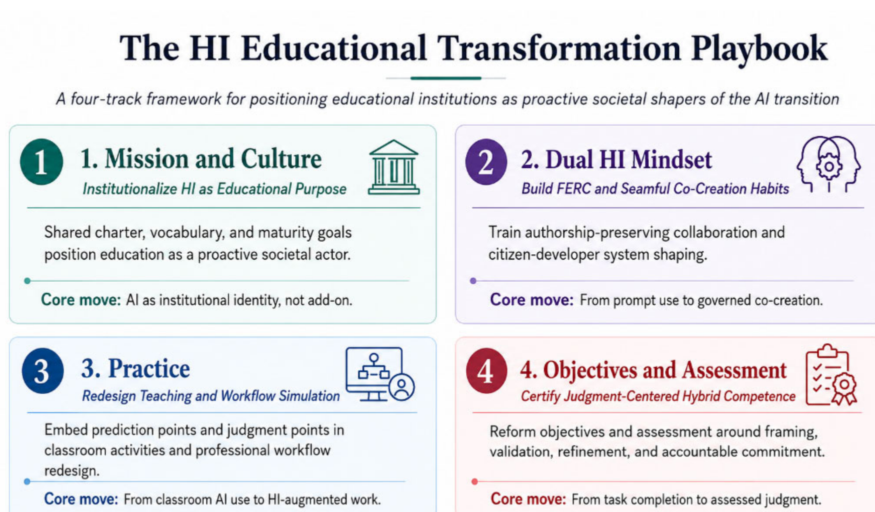
If purpose can be resolved in principle, governance is where institutions feel the dilemma daily. Educational governance has spent decades shifting from professional responsibility toward measurable accountability — from trusting teacher judgment to auditing documented output (Hopmann, 2008). Generative AI lands on this regime with peculiar force: it confounds exactly the output measures around which accountability was built. When submitted artifacts no longer certify the learning they were designed to evidence, an output-accountable institution faces an impossible position — accountable for outcomes it can no longer validly measure, while the policing apparatus that promises to restore measurability fails technically and pedagogically.

The way through follows from Section 3.3: shifting evaluative weight from output authenticity toward observable process — whether learners frame before generating, demand and compare alternatives, revise criteria, and commit with documented rationale. Graduated frameworks such as the AI Assessment Scale, which specify the permitted role of AI per task rather than issuing blanket rules, point in the same direction and are already entering institutional practice (Perkins et al., 2024). But the deeper implication should be stated plainly: process-based assessment re-professionalizes teaching. Judging the quality of a learner's framing, evaluation, and commitment is an exercise of professional judgment, not the administration of a judgment-free metric — which means the reform proposed here runs against the accountability current Hopmann describes and requires institutions to reinvest in, and re-trust, teacher judgment, with real costs in professional development and assessment redesign that the maturity model of Section 5.3 treats as a capability to be built in stages. There is also a coherence argument institutions can hardly escape: a school that preaches judgment as the human contribution while governing its own teachers by judgment-free metrics teaches the opposite of what it says. Institutional transformation is thus not only the delivery mechanism of hybrid-intelligence education but part of its content — students learn what their institutions model.

The stakes return us to Section 1. Whether education becomes a steering institution of the AI transition or remains its reactive adjustment mechanism is decided not in ministries alone but in the accumulated choices of institutions like those of Section 6 — in whether they resolve the calibration dilemma toward the change-agent mission, anchor transformation in formation as well as qualification, and accept the governance costs of assessing process. The remainder of this paper provides the operational architecture for institutions prepared to make that decision.

5. The Hybrid Intelligence Educational Transformation Playbook

The preceding sections established the rationale (Sections 1–2), the interaction architecture (Section 3), and the institutional mission (Section 4). This section translates them into an actionable reform framework. One design principle requires development first, because it carries the playbook's deepest ambition: the transition *from designing learning to learner designing*.



5.1 From Designing Learning to Learner Designing: Friction, Agency, and Seamful Co-Creation

That learning requires resistance is one of the most consistently rediscovered findings in educational research: from Dewey's "felt difficulty" as the origin of reflective thinking (Dewey, 1933), through desirable difficulties (Bjork & Bjork, 2011) and productive failure (Kapur, 2016), to the finding that learning events cluster at impasses where fluent progress breaks down (VanLehn et al., 2003). Generative AI removes this resistance indiscriminately — the mechanism behind the evidence of Section 2.2 — and a recent design literature responds by deliberately reintroducing it: AI that challenges rather than obeys, questions rather than answers, and surfaces the metacognitive demands that fluent generation conceals (Sarkar, 2024; Tankelevitch et al., 2024; Fan et al., 2025).

These interventions are valuable, but they share a structural assumption worth making explicit: design authority remains with the institution. The learner is presented with a *designed learning situation* — a curated set of approved GenAI tools, a teacher-authored bot with pedagogical guardrails, a friction mechanism that slows or interrogates, an assessment scale specifying AI's permitted role in each task (Perkins et al., 2024). The learner acts, reflects, and evaluates *inside* an environment configured elsewhere; the tools arrive as finished furniture. For foundational stages this is appropriate. But as an endpoint it undershoots both the formation argument of Section 4.4 — judgment is a capacity exercised, not administered — and the mission of Section 4.3: graduates expected to redesign professional workflows must have practiced designing, not merely inhabiting, hybrid work situations.

The HI playbook therefore prescribes a deliberate progression of design agency: from the institution-designed situation, to learner agency over the learning situation — task, workflow, and, crucially, the tool itself. What makes this pedagogically realistic now is a technological shift: generative AI has democratized system-shaping. Through no-code configuration, custom bots, prompt-defined workflows, and lightweight integrations, non-programmers act as *citizen developers*, reshaping system behavior without software training — capabilities that have not yet solidified into teaching practice and therefore constitute an open design space that learners and teachers can occupy together. AI systems stop being fixed tools designed elsewhere and become malleable infrastructure. The classroom question shifts from "which tools do we allow?" to "who shapes the tool?"

The design vocabulary for this shift comes from human–computer interaction. Against the ideal of seamless technology that disappears into use (Weiser, 1991), *seamful design* treats the seams of a system — the points where it is incomplete, uncertain, or dependent on human interpretation — as resources to be selectively revealed rather than hidden (Chalmers & Galanil, 2004). Ehsan and colleagues' *seamful XAI* operationalizes the principle for AI: deliberately crafting the junctures where system limits, uncertainty, and responsibility handovers become visible and actionable (Ehsan et al., 2024). Education has unnamed antecedents — hybrid human–AI regulation already models transitions of control between learner and system as designed handovers (Molenaar, 2022) — and in the vocabulary of this paper, a seam is simply a judgment point made structural: the scaffold-withdrawal of Section 3.2, placed where the prediction–judgment mapping of Section 2.1 requires human cognition to engage. The HI appropriation goes one step further than detection: learners and teachers should *engineer* seams — define task boundaries before delegating, configure prompts and evaluation criteria, mark handover and escalation points, document how AI contributions enter final outputs, and iteratively reshape the workflow as purposes evolve. Teachers become workflow architects; students become end-user designers of their own learning situations; and friction, rather than being abolished or imposed, is *aimed* — placed by co-creative design exactly where the desirable-difficulties tradition says the target cognition lives. The figure below summarizes the contrast.

From Friction to Seamful Co-Creation

Two approaches to structuring human–AI interaction in education

Aspect	Friction	Seamful co-creation
Core idea	Introduce pauses or resistance to slow automation and promote reflection	Make AI's limits and handover points visible to structure human–AI complementarity
Primary focus	Regulate student behavior and prevent shortcuts	Engineer explicit division of labor between prediction and judgment
Relation to technology	Technology constrained through rules and restrictions	Technology treated as malleable infrastructure to be configured and shaped
Role of teacher & student	Teacher moderates; student reflects and evaluates	Teacher designs workflows; student acts as co-creator / citizen developer
Handling uncertainty & trust	Surface ambiguity to encourage critical questioning	Define escalation points, responsibility boundaries, and verification practices
Educational trajectory	Classroom-level intervention to improve learning quality	Institutional architecture that cultivates durable hybrid intelligence

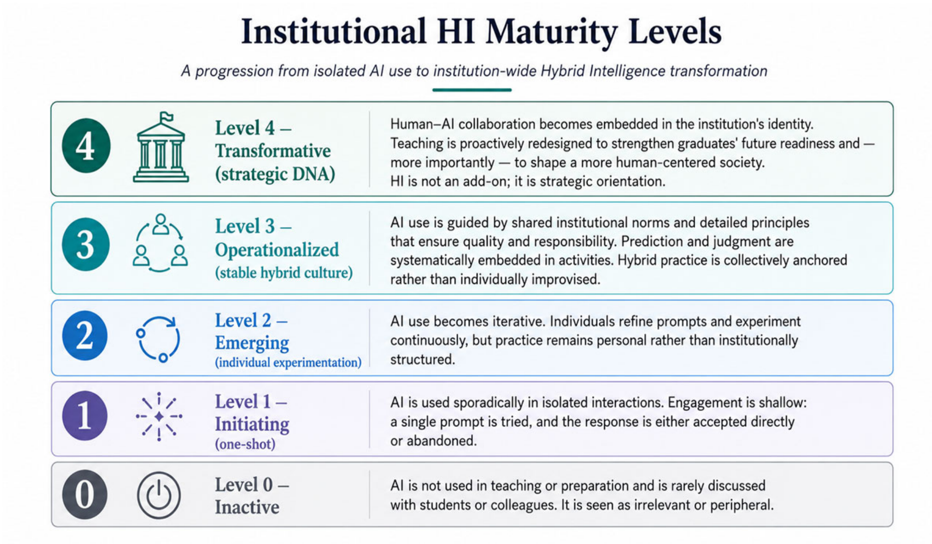
5.2 The Playbook Architecture: Four Tracks

Institutional transformation classically proceeds through four stages: redefine the **mission**, build the **mindset** that carries it, redesign daily **practice**, and rewrite the **formal structures** that reward it. The playbook follows exactly this sequence — mirroring the 4P organizational transformation model of the companion framework (Sherson et al., 2025, 2026a), so that educational institutions can borrow mature change instruments rather than improvise, and so that graduates experience daily the transformation architecture they will later carry into workplaces (Section 4.3). Track 1 recasts institutional mission and measures cultural maturity; Track 2 trains the dual HI mindset; Track 3 redesigns teaching practice and simulates professional workflow re-engineering; Track 4 rewrites learning objectives and assessment.

5.3 Track 1 — Mission and Culture: Charter, Maturity, and the Institution as Societal Actor

The first track institutionalizes the mission argued for in Section 4.3. Institutions adopt a shared HI charter — foundational principles such as: AI is a co-creative assistant, not an authority; never treat AI as a knowledge expert beyond your own domain; use FERC whenever human authorship matters; work in steps small enough to request options, not final answers. The explicit institutional goal is Level-4 maturity, *HI as institutional identity*, in which the institution continuously redesigns education around how AI and human judgment are reshaping work and educates graduates not merely to be employable but to proactively steer the future of work toward human-centered sustainability — the living-laboratory mission and the disciplinary chaos pilot of Section 4.3, given governance form.

Cultural progress is made measurable through periodic maturity assessment with a common instrument, repeated each term, adapting the organizational collaboration-maturity model (Sherson et al., 2026c) to educational settings:



These levels provide a developmental horizon and a self-diagnostic: are AI interactions structured around framing, refinement, and accountable commitment? Is hybrid competence collectively cultivated or individually improvised? Is HI treated as operational convenience — or institutional mission? Culture-level formulations track the accompanying shift in institutional self-narrative — from "AI is irrelevant or threatening" (Level 0), through "is this cheating?" (Level 1) and "AI is legitimate but skill-dependent" (Level 2), to "AI collaboration *is* professional excellence" (Level 3) and "human–AI partnership is our identity" (Level 4) — a trajectory the IBC project title, "from cheating to hybrid intelligence," captures in a single phrase (Section 6.1).

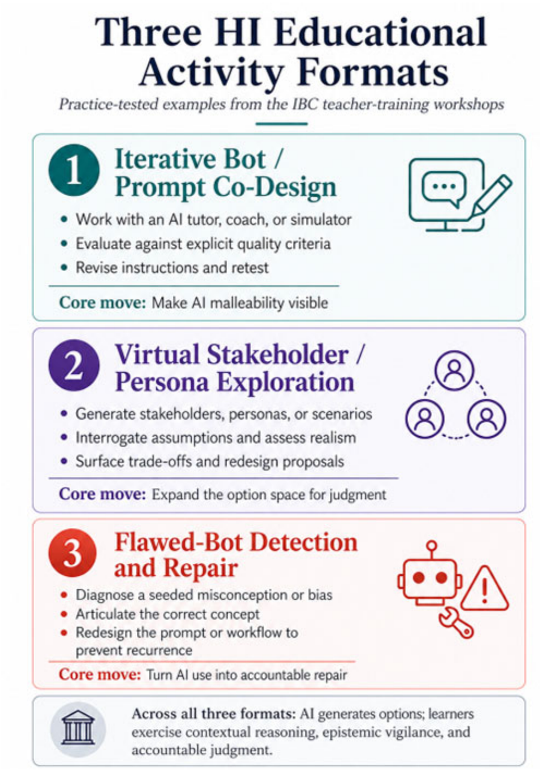
5.4 Track 2 — The Dual HI Mindset: Authorship-Preserving Collaboration and Seamful Tinkering

Culture is carried by mindset, and the playbook trains two complementary orientations — two, because they answer the two failure modes identified in Section 2: passive consumption of fluent output, and passive acceptance of fixed tools.

The first is the **FERC mindset**: collaborative but authorship-preserving. Learners and staff internalize the governed interaction discipline of Section 3 — frame before generating, demand alternatives, evaluate comparatively, commit accountably — until it operates as professional habit rather than imposed procedure. Institutional supports include baseline FERC training for all staff and students, and prompt libraries as institutional memory ("prompt of the week," shared exemplars, documented usage patterns) that convert individual craft into collective capability.

The second is the **seamful citizen-developer mindset**: exploratory, tinkering, system-shaping. Where the FERC mindset governs interaction *within* a given configuration, the seamful mindset treats the configuration itself as an object of design (Section 5.1): learners probe AI systems for seams, reconfigure, stress-test, and improve them — the mindset of playful epistemic agency that converts the "apprentice workflow designer" role (Section 2.1) into daily practice. The two mindsets are trained together because each disciplines the other: FERC without tinkering ossifies into ritual; tinkering without FERC dissolves into unaccountable play. Their union — governed authorship over exploratively shaped systems — is the operational meaning of hybrid competence.

5.5 Track 3 — Practice: Teaching Activity Redesign and Workflow Re-Engineering



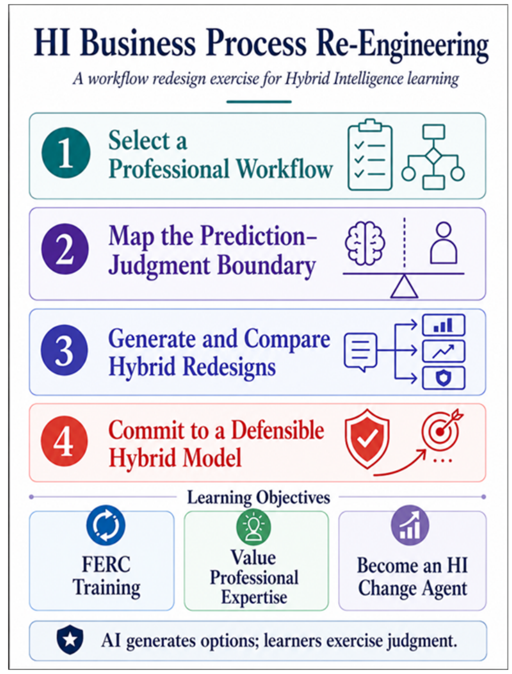
If Tracks 1 and 2 establish culture and mindset, Track 3 translates them into classroom design. Its core principle is that every learning activity should contain explicit **prediction points** and **judgment points**: moments where AI expands, accelerates, or simulates, and moments where learners frame, critique, verify, refine, and assume responsibility. Without this deliberate allocation, AI integration risks producing speed without depth.

The following exercise formats illustrate this principle. They are not an exhaustive catalogue, but practice-tested examples from the IBC teacher-training workshops (Section 6.1), where they were developed and iterated with staff across business and vocational programs. In **Iterative Bot / Prompt Co-Design**, students work with an AI configuration — such as a tutor, debate partner, case coach, or workflow simulator — evaluate it against explicit quality criteria, and revise its instructions before retesting. The exercise makes AI malleability visible: systems are not fixed authorities, but configurable components whose limitations become seams for learner redesign (Raz et al., 2024). In **Virtual Stakeholder / Persona Exploration**, AI generates stakeholders, user personas, or scenario variations that expand the analytical field; students then interrogate assumptions, assess realism, surface trade-offs, and redesign proposals accordingly. In **Flawed-Bot Detection and Repair**, an AI system is intentionally seeded with a misconception or bias, requiring students to diagnose the flaw, articulate the correct concept, and redesign the prompt or workflow to prevent recurrence. Across all three formats, AI handles generative expansion while learners exercise contextual reasoning, epistemic vigilance, and accountable judgment.

These formats can be implemented through different methods — ad hoc prompting, fixed template prompts, or custom GPTs and agents — each with trade-offs in flexibility, robustness, transparency, and learner ownership (see Appendix Table A1). Together, they form a continuum from conversational tool use to structured workflow scaffolding to learner-configured systems. As practice moves along this continuum, agency shifts from tool use toward system shaping, enacting the progression from designed learning to learner designing established in Section 5.1. The constant principle is that prompts do not create educational value; disciplined refinement does. AI outputs become meaningful only when subjected to structured human critique and accountable commitment. The classroom thereby becomes a laboratory for HI itself: a place where students learn not merely to work with AI, but to design, evaluate, and govern human–AI collaboration.

Re-imagining professional workflows: HI Business Process Re-Engineering. The track's capstone format extends HI from learning tasks to professional futures. A central ambition of the HI Manifesto is to create organizations that compete not through automation alone, but through structured complementarity between machine prediction and human judgment: predictable subtasks are automated, while human actors move toward richer evaluative, contextual, and strategic work (Sherson et al., 2026a). Educationally, this becomes a workflow exercise. Students select a concrete professional process — pricing decisions, complaint handling, audit procedures, marketing campaign design, case assessment — and redesign it by identifying which elements are prediction-heavy and AI-augmentable, and which require human judgment, accountability, and contextual sense-making.

In the IBC workshop series, this format was implemented as a structured Method-2 exercise (see Appendix Table A1): students adapted multi-step prompt sequences that guided them through mapping the current process, identifying prediction–judgment boundaries, generating alternative hybrid configurations, and exploring business-model implications. The scaffolding ensured comparability across groups while keeping attention on judgment quality rather than prompt syntax.



The exercise is not about producing impressive AI outputs. Its pedagogical core is a FERC-style experience: students frame the professional context and success criteria, explore AI-generated redesign options, refine them through theory-informed critique, and commit to a defensible hybrid configuration. The key learning occurs during refinement. Students discover that without domain expertise — legal nuance, financial risk assessment, ethical trade-off awareness, organizational constraints — AI proposals remain generic and commoditized. What initially appears innovative becomes recognizable as a standardized output stream unless shaped by expert judgment. The double bind of Section 4.2 is thus made experiential: students feel why chatbot-substitutable output commands no premium, and why accumulated domain judgment is what converts AI assistance into value.

Workflow re-imagination therefore serves both as applied learning and as motivational architecture. It consolidates students' existing competences in conceptual reasoning, contextual analysis, and evaluative comparison, while making visible the tacit, experience-based judgment they have not yet fully acquired. In doing so, it reframes the opposition between "just-in-case" and "just-in-time" education: foundational knowledge is not displaced by AI, but becomes the condition for meaningful adaptation. By using AI as a simulation engine while preserving judgment as the locus of value, HI Business Process Re-Engineering bridges classroom abstraction and real-world complexity, allowing expertise to begin accumulating before labor-market entry.

5.6 Track 4 — Redefining Learning Objectives for Hybrid Intelligence and Employability

The most decisive reform in HI education concerns the redefinition of learning objectives themselves. As long as curricula are defined around tasks that AI can increasingly perform predictively — registering, drafting, classifying, calculating, coding — education risks certifying shrinking domains of value. The central shift is therefore from task execution to judgment-centered augmentation: routine elements become AI-assisted, while educational value lies in the student's ability to frame problems, validate outputs, critique assumptions, and responsibly finalize AI-supported work. Foundational knowledge is not reduced; it becomes the basis for exercising informed judgment.

A "Trade Assistant" example illustrates the shift. Where students might traditionally learn to process customer orders correctly in an ERP system, an HI objective asks them to evaluate AI-generated order proposals, detect anomalies, and justify override or acceptance decisions in light of pricing patterns, customer history, and operational risks. Such exercises make visible where prediction ends and professional judgment begins. Without domain expertise, AI outputs remain structurally plausible but contextually thin; with expertise, AI becomes a force multiplier. The figure below generalizes this pattern across domains.

From Traditional to HI Learning Objectives

Illustrative shifts from task execution to judgment-centered augmentation

Domain	Traditional Learning Objective	HI Learning Objective
Trade & Logistics	Process customer orders correctly in ERP and ensure accurate invoicing.	Validate and critically assess AI-generated order proposals; detect anomalies and justify override decisions based on context and customer history.
Business / Marketing	Produce a complete market analysis report.	Evaluate, refine, and justify an AI-supported market analysis, identifying assumptions, risks, and strategic trade-offs.
Engineering	Calculate structural load requirements using standard formulas.	Assess and stress-test AI-assisted load calculations, identify boundary conditions, and justify final design decisions under uncertainty.
Healthcare	Develop a treatment plan based on patient symptoms and guidelines.	Critically evaluate AI-generated treatment suggestions, integrate patient-specific factors, and assume accountable clinical decision-making.

The reformulated objectives share a common structure: each is stated as observable judgment behavior — validate, detect, justify, stress-test, integrate, assume accountability — rather than task completion. This is deliberate. Judgment becomes teachable only when learners are given a structure for exercising it; otherwise, judgment-centered tasks can feel arbitrary, especially when AI outputs are fluent but shallow. The role of FERC, the two-alternatives rule, and the seam vocabulary is precisely to convert "exercise judgment" from an exhortation into a teachable procedure.

Implementation follows three steps: identify objectives likely to become prediction-heavy; reformulate them as HI objectives emphasizing judgment; and attach at least one Track-3 exercise that trains the reformulated objective. Assessment then follows the objectives rather than fighting them: evaluative weight shifts from output authenticity toward observable process — documented FERC behavior, graduated specification of AI's permitted role per task (Perkins et al., 2024), and justification of seam decisions. Because the objectives are phrased as observable behaviors, constructive alignment is direct: objective, learning activity, and assessment become the same object viewed three ways (Biggs, 1996).

The employability implication completes the playbook. HI objectives describe competences that are legible to employers because they mirror how value is created in AI-rich workplaces: not through the execution AI commoditizes, but through the framing, validation, and accountable finalization that the human-touch marketplace prices. A graduate certified against such objectives carries evidence of what the entry-level labor market increasingly lacks: assessed experience exercising judgment over AI-supported work. Redefining learning objectives is therefore where the institutional transformation and the economic argument meet — the point at which education stops certifying what AI has already absorbed and starts certifying what only its graduates can supply.

6. Early Institutional Realizations: Two Cases

A framework that claims to be an operating model rather than an aspiration must ultimately be tested in institutional practice. While systematic evaluation lies beyond the scope of this paper, the playbook presented in Section 5 is already being enacted in one full institutional deployment and designed into a second, deliberately contrasting setting. We present these as formative cases in the tradition of educational design-based research, where frameworks are developed through iterative implementation in authentic contexts rather than validated only after the fact (Design-Based Research Collective, 2003; Anderson & Shattuck, 2012). The cases document institutional design decisions and transformation structures, not learning outcomes; both remain in progress, and detailed process documentation and evaluation will be reported elsewhere.

The cases were selected for maximal contrast rather than convenience (cf. Flyvbjerg, 2006). IBC represents an institution-wide, leadership-anchored transformation in upper-secondary business and vocational education, focused primarily on digital knowledge work. ETH Student Project House represents a bottom-up, community-driven extension into a university maker-space context, where physical making and prototyping foreground tacit, embodied knowledge. If the framework's core elements prove workable across these contrasting settings, its broader applicability is strengthened; where they diverge, they reveal genuine boundary conditions. A fuller comparison is provided in Appendix Table A2.

6.1 IBC: Institution-Wide Transformation in Vocational and Business Education

The first case is IBC (International Business College), one of Denmark's largest business colleges. Its project **"AI — fra snyd til hybrid intelligens"** ("AI — from cheating to hybrid intelligence"), funded by Region Syddanmark's education fund, develops a generative-AI strategy for the HHX and EUD/EUX programs across the college's campuses (Region Syddanmark, 2024). The title captures the reframing ambition: to move AI from an integrity threat to be policed toward a governed instrument of student learning — the system-level shift this paper argues education must undertake.

The project's institutional standpoint mirrors the diagnosis developed here. IBC holds that the school's dual purpose — subject competence on one side, personal formation (*dannelse*) and autonomy on the other — remains unchanged, but that its traditional means have been destabilized: written and productive work once used to elicit reflection and critical competence can now be generated as output without the learning process it was designed to support. The standpoint also identifies the risk that unguided AI access widens the competence span between students who use AI to evade learning and those who use it to seek deeper learning.

The transformation is organized around three tracks — technology, didactics, and formation — launched through a cross-campus kickoff and developed in department-level action-learning groups. Parallel strands include teacher professional development, shared rules of engagement for AI use, a common framing model for specifying AI's role in tasks, experimentation with teacher-configured AI assistants, and a shared vocabulary for structured human–AI interaction. In this form, the project instantiates the playbook's core sequence: institutional positioning, capability building, shared interaction norms, and iterative local experimentation. Notably, the kickoff synthesis converged on a central principle of this paper: learning and formation require friction, and uncritical AI use risks hollowing out precisely that process.

A positionality note is required: the first author serves on the project's advisory board. This role is compatible with design-based research, where the researcher often participates in design, but it precludes treating the case as independent evaluation (Anderson & Shattuck, 2012). Detailed results from the project, including the evolution of its rules of engagement and departmental action-learning cycles, will be reported elsewhere; independent outcome studies remain essential future work.

6.2 ETH Student Project House: Extending Hybrid Intelligence into Physical Making

The second case, in planning at the time of writing, tests the framework in a domain largely absent from the AI-in-education debate: physical making. ETH Student Project House (SPH) is ETH Zürich's cross-disciplinary maker space, where students define their own challenges and build physical and digital prototypes through trainings, coaching, and community formats.

The planned collaboration has two levels. First, a dedicated workshop will deep-test the framework. Participants begin with a macro-FERC cycle, using the Double-FERC interpretation of design (Section 3.1; Sherson et al., 2026c), to construct their project challenge before building begins and to anchor it in grand-challenge framings. During construction, mini-FERC checkpoints introduce periodic pauses in which participants document the artifact-in-progress, reflect on what it reveals against the original frame, generate alternative next steps with AI support, and commit deliberately to the next construction move. This checkpoint rhythm extends FERC and seamful design into embodied work, where standard digital interaction guidance applies only partly.

Second, contingent on workshop testing, the HI mindset and FERC discipline will be integrated into SPH's broader portfolio of trainings, coaching interactions, and community formats, moving from bounded experiment to shared facilitation layer. This mirrors the playbook's maturity logic: governed testing in a protected setting first, institutional embedding as shared practice second.

The case is deliberately exploratory. Physical making foregrounds tacit, embodied knowledge — the far side of the prediction–judgment frontier — and therefore tests whether HI-structured interaction can support domains where much of the relevant knowledge is slow to externalize. If the checkpoint rhythm supports such work, the framework's generality is strengthened; if it proves intrusive or ritualistic, that boundary condition is equally informative. The collaboration is also designed as a co-evolution experiment: formats will be iterated with the student community, treating learners as co-designers of hybrid practice rather than recipients of it.

Together, the two cases show that the playbook can be enacted within real institutional constraints and adapted across contrasting educational levels, change directions, and task domains. They also illustrate the design pressure generated by formative implementation: in the IBC case, for example, practitioners' emphasis on friction and resistance fed back into the seamful design principle developed in Section 5. These cases do not establish learning outcomes, comparative effectiveness, or durability; those claims require independent, controlled, and longitudinal studies, as discussed in Section 7. In design-based-research terms, the framework is completing its first design cycles; the contribution of this paper is the theoretically grounded architecture those cycles are testing.

7. Discussion and Conclusion

This paper has developed the educational pillar of the Hybrid Intelligence research program. It has argued that education must transform not only because AI changes teaching and learning, but because it changes the conditions under which human judgment remains socially and economically valuable. The paper makes four contributions. First, it grounds HI education in an explicit economic rationale: the dynamic prediction–judgment frontier, the expertise-pipeline coordination failure, and the human-touch marketplace explain why judgment-centered education is structurally necessary, not merely pedagogically desirable. Second, it replaces static catalogues of "uniquely human" capacities with an operational account of complementarity based on the knowledge structure of tasks. Third, it presents FERC as a pedagogically grounded operating model for governed human–AI collaboration, making framing, exploration, refinement, and commitment observable, teachable, and assessable. Fourth, it translates these foundations into an institutional transformation playbook through which education can move from ad hoc AI experimentation toward societal stewardship of the AI transition.

Across these contributions runs a single design philosophy: **operationalization as democratization**. The prediction–judgment test, the FERC cycle, the two-alternatives rule, the seam vocabulary, and the maturity instruments each convert what might otherwise remain an expert judgment call or hoped-for disposition into a procedure that learners can practice, teachers can assess, and institutions can govern. If hybrid competence is to become a general educational outcome rather than the craft of the already metacognitively skilled, this conversion is the mechanism of scale.

The framework remains prescriptive and formative rather than outcome-validated. The cases presented here document implementability and design-cycle learning, not learning gains, comparative effectiveness, or durability; the first author's advisory role also precludes treating them as independent evaluations. FERC itself raises the probability of genuine self-regulation without guaranteeing it: learners can traverse its phases while tacitly adopting AI-suggested framings and performing authorship without enacting it. Scope limits also remain. FERC is designed primarily for solution-oriented tasks; the exercise repertoire is drawn mainly from business, vocational, and technical education in a Nordic institutional context; and transfer to other disciplines, age groups, and governance cultures remains to be demonstrated. Finally, the equity challenge is substantial: seamful, judgment-centered education presupposes access to configurable AI tools and trained teachers, and without deliberate policy attention could widen the competence gaps it aims to close.

These limitations define the research and action agenda. Priority work includes controlled and longitudinal studies comparing FERC-structured with unstructured AI-supported learning on durable learning and transfer; instrumented interaction analysis linking observable phase behavior to learning outcomes; and longitudinal tracking of graduates into AI-rich workplaces to test whether HI objectives improve employability and workflow-design capacity. Assessment research must also move beyond AI literacy toward hybrid performance: whether learner–AI combinations outperform either alone, and whether that superiority grows with training. The ETH collaboration will probe boundary conditions in embodied, tacit-knowledge-intensive domains, while extensions to early education, humanities disciplines, and low-resource contexts are needed to identify where the architecture holds and where it must adapt. We invite educational institutions, researchers, and policymakers to adopt, adapt, and stress-test the playbook through the Grand Hybrid Intelligence Challenges coordination mechanism hosted at the HI Manifesto website.

The central claim is ultimately institutional. Education faces the AI transition holding more leverage than the current debate assigns it. The prevailing framings — policing integrity, teaching literacy, awaiting regulation — cast schools and universities as sites where a technological wave must be managed. This paper argues instead that education is one of the institutions on which the direction of that wave depends. The AI economy will become automation-dominant or genuinely hybrid depending in part on whether human judgment remains structurally central to value creation; human judgment remains central only if it is reliably reproduced; and its reproduction, increasingly underprovided by firms responding rationally to short-term incentives, now falls to education as a defining societal task.

That task is achievable with instruments education already understands: an interaction discipline with a century of learning theory behind it, a design principle that aims friction where learning lives, a transformation architecture borrowed from institutional change practice, and an assessment logic that re-professionalizes teaching. It is also a task worth wanting, not only needing. The capacities a sensecurity society demands — framing, judgment, accountable commitment, and the ability to redesign human–AI workflows — are also capacities the formation tradition has long recognized as central to autonomy and responsible participation. In the age of abundant prediction, education serves both its economy and its ideals by institutionalizing society's capacity to shape the AI transition. Its graduates, disciplinary chaos pilots in the making, can then help build the hybrid future rather than merely undergo the automated one.

References

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3–30.

Acemoglu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker AI? Choosing a path of machines in service of minds. MIT Shaping the Future of Work Initiative, policy memo. (Cite if you keep the pro-worker framing anywhere in the education paper; the macro paper builds on it.)

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction, judgment, and complexity: A theory of decision-making and artificial intelligence. In The economics of artificial intelligence: An agenda (pp. 89–110). University of Chicago Press.

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. Information Economics and Policy, 47, 1–6.

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., ... & Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer, 53(8), 18–28.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? Educational Researcher, 41(1), 16–25.

Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis (OECD Social, Employment and Migration Working Papers No. 189). OECD Publishing.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3–30.

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

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. Educational Assessment, Evaluation and Accountability, 21(1), 33–46.

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher et al. (Eds.), Psychology and the real world (pp. 56–64). Worth.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies: "Engines of growth"? Journal of Econometrics, 65(1), 83–108.

Brynjolfsson, E., Chandar, B., & Chen, R. (2025). Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence (Working Paper No. 25-027). Stanford Institute for Economic Policy Research. https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence/

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. Quarterly Journal of Economics, 140(2), 889–942.

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333–372.

Chalmers, M., & Galani, A. (2004). Seamful interweaving: Heterogeneity in the theory and design of interactive systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 243–252). ACM. https://doi.org/10.1145/1013115.1013149

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453–494). Erlbaum.

Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European Digital Competence Framework (5th ed.). Publications Office of the European Union.

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. British Journal of Educational Technology, 56(2), 469–488.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. American Economic Review, 80(2), 355–361.

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality* (Harvard Business School Technology & Operations Management Unit Working Paper No. 24-013). Harvard Business School. https://doi.org/10.2139/ssrn.4573321

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. Business & Information Systems Engineering, 61(5), 637–643.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, 32(1), 5–8.

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. D.C. Heath. (Original work published 1910)

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290.

Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. Free Press.

Ehsan, U., Liao, Q. V., Passi, S., Riedl, M. O., & Daumé III, H. (2024). Seamful XAI: Operationalizing seamful design in explainable AI. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW). [verify volume/pages]

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Science, 384(6702), 1306–1308.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & 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.

Finke, R. A. (1996). Imagery, creativity, and emergent structure. Consciousness and Cognition, 5(3), 381–393.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative Inquiry, 12(2), 219–245.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6.

Getzels, J. W., & Csikszentmihalyi, M. (1976). The creative vision: A longitudinal study of problem finding in art. Wiley.

Giannakos, M. (2025, October). Hybrid intelligence and human learning. In Proceedings of the 36th Annual Conference of the European Association of Cognitive Ergonomics (EACE) (pp. 1–2).

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering, 51*, 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, 105681.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542–570.

Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. Journal of Curriculum Studies, 40(4), 417–456.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. Organizational Behavior and Human Decision Processes, 67(3), 247–257.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. Educational Psychologist, 48(1), 25–39.

Järvelä, S., Zhao, G., Nguyen, A., & Chen, H. (2025). Hybrid intelligence: Human–AI coevolution and learning. *British Journal of Educational Technology, 56*(2), 455–468. https://doi.org/10.1111/bjet.13560

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67(1), 5–32.

Lodefalk, M., Löthman, L., Koch, M., & Engberg, E. (2026). *Same storm, different boats: Generative AI and the age gradient in hiring* (Working Paper No. 2/2026). Örebro University School of Business. https://hdl.handle.net/10419/339320

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51(2), 289–299.

Kim, Y., Lee, J., Kim, S., Park, J., & Kim, J. (2024). Understanding users' dissatisfaction with ChatGPT responses. In Proceedings of the 29th International Conference on Intelligent User Interfaces (pp. 385–404). ACM.

Klafki, W. (2000). Didaktik analysis as the core of preparation of instruction. In I. Westbury, S. Hopmann, & K. Riquarts (Eds.), *Teaching as a reflective practice: The German Didaktik tradition* (pp. 139–159). Lawrence Erlbaum Associates.

Kleinmintz, O. M., Ivancovsky, T., & Shamay-Tsoory, S. G. (2019). The two-fold model of creativity: The neural underpinnings of the generation and evaluation of creative ideas. *Current Opinion in Behavioral Sciences, 27*, 131–138. https://doi.org/10.1016/j.cobeha.2018.11.004

Lebuda, I., & Benedek, M. (2023). A systematic framework of creative metacognition. Physics of Life Reviews, 46, 161–181.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM.

Madsen, P. K. (2004). The Danish model of "flexicurity": Experiences and lessons. Transfer: European Review of Labour and Research, 10(2), 187–207.

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., ... & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? Computers and Education: Artificial Intelligence, 3, 100056.

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO.

Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO.

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. Creativity Research Journal, 4(2), 91–122.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education, 49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology, 21*, 34–77. https://doi.org/10.1080/10463281003765323

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192.

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

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(6). https://doi.org/10.53761/q3azde36

Polanyi, M. (1966). The tacit dimension. University of Chicago Press.

Rafner, J. F., Dellermann, D., Hjorth, A., Verasztó, D., Kampf, C. E., et al. (2021). Deskilling, upskilling, and reskilling: A case for hybrid intelligence. *Morals & Machines, 1*(2), 24–39. https://doi.org/10.5771/2747-5174-2021-2-24

Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., & Sherson, J. (2023). Creativity in the age of generative AI. Nature Human Behaviour, 7, 1836–1838.

Raz, T., Reiter-Palmon, R., & Kenett, Y. N. (2025). The role of asking more complex questions in creative thinking. *Psychology of Aesthetics, Creativity, and the Arts, 19*(6), 1505–1525. https://doi.org/10.1037/aca0000658

Region Syddanmark. (2024). AI — fra snyd til hybrid intelligens https://regionsyddanmark.dk/uddannelse-trafik-og-samarbejde/uddannelse/uddannelsespuljen/projektbanken/digitalisering-ai/ai-fra-snyd-til-hybrid-intelligens

Reiter-Palmon, R., & Robinson, E. J. (2009). Problem identification and construction: What do we know, what is the future? Psychology of Aesthetics, Creativity, and the Arts, 3(1), 43–47.

Sarkar, A. (2024). AI should challenge, not obey. Communications of the ACM. [verify issue/pages]

Sherson, J., Rafner, J., Bang, J. B., … Tiddi, I., Tudoran, A. A., & Wagner, M. (2026a). The Hybrid Intelligence Manifesto: A multi-level socio-technical framework for keeping human judgement economically valuable. HI Infinity. https://hi-infinity.ai/research/the-hybrid-intelligence-manifesto/

Sherson, J., Rafner, J., Reiter-Palmon, R., Lebuda, I., Söllner, M., Kenett, Y. N., Emruli, B., Rindle, M., Weiss, S., Goecke, B., Vinchon, F., Mao, Y., Dellermann, D., Rahimim, S., Marler, J. H., Nguyen, A., Bjerring, J. C., & DiPaola, S. (2026c). Ensuring human authorship in the age of generative AI: The FERC framework for governed human–AI collaboration. In Proceedings of the 14th Nordic Conference on Human–Computer Interaction (NordiCHI 2026). ACM. https://hi-infinity.ai/research/ensuring-human-authorship-in-the-age-of-generative-ai/

Sherson, J., Schroeder, S., Pytlikova, M., Rafner, J., Lodefalk, M., Gordon, A., Rahimi, S., Söllner, M., Wagner, M., & Holm, A. B. (2026b). The economics of hybrid intelligence. HI Infinity. https://hi-infinity.ai/research/the-economics-of-hybrid-intelligence/

Sowden, P., Pringle, A., & Gabora, L. (2018). The shifting sands of creative thinking: Connections to dual-process theory. In K. J. Gilhooly, L. Ball, & L. Macchi (Eds.), *Insight and creativity in problem solving* (pp. 40–60). Routledge. https://doi.org/10.4324/9781315144061-3

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. ACM. (Note: Lev Tankelevitch is a co-author of the HI Manifesto — worth flagging internally.)

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? In Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 9795–9799.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? Cognition and Instruction, 21(3), 209–249.

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

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19(1), 26.

Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), 94–104.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. arXiv. https://arxiv.org/abs/2302.11382

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 39.

Zielińska, A., Lebuda, I., Czerwonka, M., & Karwowski, M. (2025). Self-regulation prompts improve creative performance. The Journal of Creative Behavior, 59(1), e674.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70.

Appendix

Table A1.

Three Implementation methods for HI exercise formats

From the IBC teacher-training workshops

	1) Instruction & prompting (ad hoc in chat)	2) Prompt copy-pasting (fixed template prompts)	3) Custom GPT / agent
Short description	The teacher gives a verbal/written instruction and participants write their own prompts from scratch in a standard GPT/Copilot chat.	The teacher provides 1–3 tested prompts that participants copy and adapt by filling in fields (e.g., [organization], [case], [goal]).	The teacher builds an agent with fixed instructions, tone, process requirements, and optionally uploaded documents/data and/or hidden rules (e.g., a Flawed Bot).
Advantages	• Maximum flexibility: adaptable on the spot to subject, level, and participants' ideas. • Good for learning to prompt as a competence (trains FERC). • Clear human ownership: participants must formulate context and criteria themselves.	• Fast and easy workflow: everyone starts immediately. • More uniform outputs → better comparability across groups. • Creates space for judgment: time is spent evaluating and improving output rather than struggling with prompts. • Well suited to recurring classroom exercises.	• High robustness and consistency. • Can be locked to syllabus/documents → fewer hallucinations, stronger academic alignment. • Enables multi-stage workflows with human decision gates (realistic HI training). • Can embed hidden misconceptions for diagnostic exercises. • Scalable across the organization.
Disadvantages	• High variation in quality: some get strong results, others get stuck. • Takes longer, as everyone must invent the prompt. • Risk of 'one-shot' use without reflection if participants cannot structure iteration.	• Less training in independent prompting (can become 'form filling'). • Can feel rigid if the case deviates from the template. • Risk that participants accept output too quickly because the prompt 'looks official.'	• Requires preparation and maintenance. • Less transparent how results are produced ('black box' risk). • Less flexible in the moment. • Dependent on platform access and IT setup.

Table A2.

Comparing the Two Selected HI Educational Transformation Cases

Contrasting institutional settings used to test the HI educational transformation framework

Aspect	Case 1: IBC (Denmark)	Case 2: ETH Student Project House (Switzerland)
Educational level	Upper-secondary business (HHX) and vocational (EUD/EUX)	University; cross-disciplinary maker space
Direction of change	Top-down: institution-wide strategy process	Bottom-up: community-driven experimentation
Task domain	Digital knowledge work	Physical making and prototyping
Transformation structure	Three-track program (technology, didactics, formation) with department-level action learning	Two-level adoption: deep workshop testing, then portfolio-wide facilitation layer
Role for the framework	Validation at full institutional scale	Boundary extension into embodied work
Status	Implementation in progress	In planning