

POSITION PAPER · EDUCATIONAL AI THEORY

*AI in Education Is **Constitutively** Different*

Two arguments — that AI in education cannot exist beyond human intelligence, and that an FDA-equivalent clearinghouse for AI-enabled educational resources is a logical necessity — presented as a complete position.

CONTENTS

Essay I · AI in Education Is Constitutively Different

Essay II · The Physician Analogy Is Incomplete Without the FDA

Xiangen Hu

Chair Professor, Department of Applied Social Sciences

Director, Educational Research Centre

The Hong Kong Polytechnic University

Human Agency Augmentation (HAA) Framework · April 2026



PART I · ENGLISH VERSION

AI in Education Is Constitutively Different

ENGLISH POSITION PAPER

FOUNDATIONAL ARGUMENT

AI in Education Is Constitutively Different

Why artificial intelligence in the classroom cannot exist independently of human intelligence — and why educators must be sophisticated consumers of AI, not its developers.

Xianguen Hu · Department of Applied Social Sciences, The Hong Kong Polytechnic University · Educational Research Centre

Artificial intelligence has penetrated nearly every consequential domain of human activity. Yet education stands apart — not merely in degree, but in kind. The distinction is not rhetorical. It is ontological. It changes what AI is allowed to be, what educators are obligated to know, and what society must demand before any AI system is permitted inside a learning relationship.

Part One · The Ontological Claim

AI in Education Cannot Exist Beyond Human Intelligence

In virtually every field where AI has proven transformative — protein folding, financial arbitrage, autonomous navigation, radiological diagnosis — the AI system operates on a well-defined input/output structure. The domain object is external to the minds of its users. A folded protein exists independently of any human understanding it. A mispriced derivative exists regardless of whether a trader grasps it. **The domain is, in a deep sense, inert with respect to human cognition:** it does not require a mind to learn it in order to be what it is.

Education is the only domain in which the *subject matter itself is the development of human intelligence*. This is not a design choice. It is the definition of the enterprise. There is no learning without a learner. There is no teaching without the transformation of a mind. Strip out the human intelligence — of both educator and student — and you have not merely a degraded educational experience. You have no educational experience at all. You have only information transfer, which is no more education than a bookshelf is a teacher.

*In every other domain, AI can exist and generate value without a human intelligence in the loop at execution time. In education, human intelligence is not in the loop — it **is** the loop. Remove it, and the loop ceases to exist.*

THE CONSTITUTIVE DISTINCTION

This distinction has a precise theoretical lineage. Vygotsky's Zone of Proximal Development (ZPD) establishes that learning occurs only in the cognitive space between what a learner can do alone and what she can do with a more knowledgeable other — an irreducibly relational, dyadic structure. Dewey's transactional theory of experience insists that education is not transmission of content from one vessel to another, but the co-construction of meaning through shared inquiry. Clark and Chalmers' extended mind hypothesis, applied to education, suggests that AI can become genuine cognitive scaffolding — but *only* as an extension of existing human intelligence, not as a replacement for it.

The implication is radical. When an autonomous vehicle navigates a highway, there is no human intelligence whose extension it must be. When an AI diagnoses a lung nodule from a CT scan, the radiologist may be absent at the moment of inference. The AI is, in those contexts, **self-sufficient with respect to its domain object**. But when an AI tutors a student, the student's mind is simultaneously the input, the process, and the output. The AI has no educational object to act upon except that mind. Remove the human intelligence, and the AI is tutoring nothing.

THE HAA TYPOLOGY — A FRAMEWORK FOR CLASSIFICATION

Within the Human Agency Augmentation (HAA) framework, AI deployments may be classified as **Type A** (replacing human agency), **Type B** (assisting human agency in discrete tasks), or **Type C** (extending human agency in ways that amplify cognitive reach and self-determination). In most domains, all three types are legitimate in appropriate contexts.

In education, **Type A deployment is not merely suboptimal — it is a categorical error**. An AI that replaces educator judgment, or treats the student as a passive recipient of algorithmically optimized content, has misunderstood what education is. Only Type C — AI as extension of the educator-student relationship — is constitutively appropriate. This is not a policy preference. It is a logical consequence of what education means.

DIMENSION	AI IN OTHER DOMAINS	AI IN EDUCATION
Domain object	External to human cognition (proteins, prices, pixels)	Human cognition itself — the learner's developing mind
Can AI act independently?	✓ Yes — value generated without human in loop	✗ No — without educator/student, no educational act exists
Relationship to human intelligence	Optional augmentation or replacement	Constitutively required — AI is the extension
Failure mode if AI acts alone	Suboptimal output, correctable	Categorical failure — the educational relationship dissolves
Appropriate HAA type	A, B, or C depending on context	Type C only — extension, never replacement
Primary accountability	Engineers, domain specialists	Educators, who remain epistemically responsible for learners

Educators Must Be Informed Consumers, Not Developers

A second error has quietly taken root in the culture of educational technology: the idea that educators should not only adopt AI tools, but *build them*. The rise of low-code and no-code AI platforms, the democratization of large language models, and the enthusiasm of an education community eager to innovate have combined to produce a dangerous illusion — that pedagogical expertise is sufficient qualification for AI development in educational contexts.

It is not. And the history of medicine offers us a clarifying analogy.

The Physician Analogy

Physicians do not synthesize pharmaceuticals. They do not design clinical trials or establish bioavailability. These tasks belong to pharmacologists, chemists, and biostatisticians — specialists whose formation is devoted to producing reliable evidence about what a compound does to a human body.

What a physician does instead is something far harder and far more important: **she applies that evidence to the irreducibly particular situation of a patient who stands before her**. She knows enough about pharmacology to evaluate a study's methodology. She knows when to prescribe, when to withhold, and when to refer. Her expertise is judgment, not synthesis.

<p>MEDICINE</p> <p>The Physician's Role</p> <ul style="list-style-type: none"> → Understands pharmacological evidence → Evaluates RCTs and systematic reviews → Makes prescription decisions for patients → Does not synthesize drugs → Does not run clinical trials → Holds epistemic responsibility for patient outcomes 	≡	<p>EDUCATION</p> <p>The Educator's Role</p> <ul style="list-style-type: none"> → Understands learning science evidence → Evaluates efficacy studies and IRT/psychometric validity → Makes pedagogical decisions for learners → Does develop AI systems → Does run their own validation studies → Holds epistemic responsibility for learner outcomes
--	---	--

The current edtech moment is, in medical terms, equivalent to encouraging every clinician to open a home chemistry lab and synthesize their own experimental compounds — then administer them to patients on the basis of the clinician's enthusiasm and anecdotal impression that "it seems to work."

Imagine if medical schools encouraged physicians to synthesize their own drugs and administer them without clinical trials, on the grounds that physicians know patients best. We would call this dangerous. We should call its educational equivalent the same thing.

THE PROFESSIONAL ANALOGY

Why Educator-Built AI Is Epistemically Perilous

Enthusiasm bias and the novelty effect. Almost any new intervention produces short-term gains — not because it is effective, but because novelty itself activates attention and motivation. Educators who develop and deploy their own AI tools are simultaneously developers, users, and evaluators — a triple conflict of interest that makes objective efficacy assessment nearly impossible.

Absence of psychometric foundations. Valid educational AI requires grounding in item response theory, learner characteristic curves, and knowledge space theory. An educator who builds an adaptive learning tool without psychometric grounding is building a tool that simulates adaptation while systematically biasing toward learners who already conform to her intuitive model.

Ethical and developmental validity gaps. AI systems that interact with learners carry risks that are not merely technical but developmental and ethical — cognitive load, motiv-

ational architecture, stereotype reinforcement, long-term metacognitive development. An educator's knowledge of her students does not, by itself, equip her to anticipate the second-order effects of a conversational AI on developing metacognition.

The irreproducibility problem. Self-developed educator tools cannot be independently evaluated, replicated, or scaled with fidelity. The evidentiary standards of educational research — pre-registration, control conditions, effect size estimation — require institutional infrastructure, methodological training, and adversarial peer review. Tools developed outside this infrastructure produce, at best, anecdote.

Efficacy Before Deployment: A New Standard

The pharmaceutical industry operates under an irreversible institutional norm: **no drug reaches patients without a structured evidence threshold**. The norm itself — that evidence precedes deployment — is not contested. No equivalent norm governs educational AI. Tools are released, marketed, adopted, and scaled at the speed of a software industry, carrying none of software's tolerance for iterative failure.

Software bugs crash programs. Educational AI bugs shape cognition, distort self-efficacy, and can generate lasting disengagement. The asymmetry between deployment speed and developmental impact is a civilizational risk that the field has not yet reckoned with.

The standard we require includes: **theoretical grounding first** (every system must articulate its theory of learning); **psychometric validity as prerequisite** (adaptive systems must demonstrate valid learner models); **efficacy evidence before scale** (pilot studies with comparison conditions and pre-registration); and **ongoing surveillance after deployment** (as in pharmacovigilance, monitoring for unexpected effects on equity, engagement, and the educator-student relationship).

AI in education is different in kind from AI in every other domain. This difference follows from the most basic fact about education: that its subject matter is the human mind in the act of becoming more fully itself.

Because of this, AI in education cannot exist beyond human intelligence. It has no object to act upon except the developing mind, and the developing mind requires a human educator — intentional, knowledgeable, ethically accountable — to remain at the center of the learning relationship. AI extends that relationship. It does not replace it.

Because of this, educators are not AI developers. They are, or must become, sophisticated and demanding consumers of AI — capable of evaluating evidence, recognizing red flags, and holding AI systems to the same standard of validated efficacy that we demand of every other intervention that operates on human health and development.

The physician does not make medicine. She knows when to prescribe it, when to withhold it, and when it is doing harm. That is the model. That is the standard. **Education must now have the courage to adopt it.**



PART III · ENGLISH VERSION

***The Physician Analogy
Is Incomplete Without the
FDA***

A PROPOSAL FOR AN EVIDENCE CLEARINGHOUSE FOR AI-ENABLED
EDUCATIONAL RESOURCES

INSTITUTIONAL COROLLARY

The Physician Analogy Is Incomplete Without the FDA

The argument that educators should be consumers, not developers of AI, is only coherent if a rigorous validation infrastructure exists. We need an FDA-equivalent and a What Works Clearinghouse for AI-enabled educational resources — and neither existing institution is sufficient.

Xianguan Hu · Department of Applied Social Sciences, The Hong Kong Polytechnic University ·
Educational Research Centre

The previous argument established two claims: that AI in education cannot exist beyond human intelligence, and that educators must therefore be sophisticated consumers of AI rather than its developers. Both claims are true. But together, they generate an obligation that has not yet been met.

The physician analogy, on reflection, contains a hidden premise. We can ask a physician to be a responsible prescriber — to evaluate evidence, exercise judgment, and protect the patient — *only because the FDA exists*. The physician's competence as a consumer of pharmaceutical knowledge presupposes an institutional infrastructure that has already done the hard work: that has evaluated safety and efficacy under controlled conditions, that has established evidence thresholds, that has the authority to withdraw products that cause harm, and that makes its findings legible to practitioners. Without the FDA, asking physicians to "just be good consumers" is not a safeguard. It is an abdication.

We cannot tell educators to be responsible consumers of AI unless we build the infrastructure that makes responsible consumption possible. The physician analogy demands the FDA. The educator analogy demands its equivalent.

THE INSTITUTIONAL COROLLARY

Education has no such infrastructure for AI-enabled tools. The What Works Clearinghouse (WWC), operated by the Institute of Education Sciences (IES), is the closest existing ana-

log — and it is valuable, rigorous, and necessary. But it was designed for a pre-AI evidence landscape. It evaluates structured curricula and discrete interventions using randomized controlled trials and quasi-experimental designs. It was not designed to evaluate adaptive systems that change their behavior in response to learner state, large language models whose outputs are stochastic and context-dependent, or AI platforms whose underlying models can be silently updated after deployment. The gap between what the WWC does and what an AI-era clearinghouse must do is not a gap of degree. It is a gap of kind.

What the WWC Was Not Built to Handle

The WWC's evidence standards are among the most rigorous in educational research. Its tiered framework — strong evidence, moderate evidence, promising evidence, no evidence — has done more to discipline educational decision-making than any other single institutional mechanism. It deserves its authority. The argument here is not against the WWC. It is that AI-enabled educational tools present a categorically different set of validation challenges that the WWC's current methodology cannot address.

The Stationarity Problem

Classical educational interventions are stationary: a reading curriculum administered in 2023 is the same curriculum administered in 2024. Its properties can be measured, and those measurements remain valid over time. An AI-enabled educational tool backed by a large language model is not stationary. The model can be updated, fine-tuned, or replaced by the vendor — silently, without notification to schools or regulators — changing the system's behavior in ways that may invalidate any prior efficacy evidence. A pharmaceutical drug approved in 2023 cannot spontaneously change its molecular structure in 2024. An AI tutoring system can, and routinely does. **Efficacy evidence for a living system requires a versioning and surveillance framework that educational research has never needed before.**

The Adaptive Validity Problem

Self-described "adaptive" educational AI systems are proliferating rapidly. Nearly every major edtech platform now claims to personalize instruction based on learner state. But adaptation is only educationally valid if the underlying learner model is valid — if the system's inferences about what a learner knows, can do, and needs next are grounded in defensible psychometric theory rather than behavioral proxies, engagement metrics, or clickstream patterns. The WWC's current framework has no mechanism for evaluating

the validity of a learner model. It can evaluate whether students who used a system outperformed those who did not. It cannot evaluate whether the system's adaptation logic was responsible for that outcome, or whether it would generalize to learners whose behavioral profile differs from the trial sample.

The Black-Box Transparency Problem

Many AI educational systems — particularly those built on proprietary large language models — operate as black boxes. Their recommendation logic, content generation parameters, and feedback mechanisms are not inspectable by educators, researchers, or regulators. This is not merely an inconvenience for researchers. It is an epistemic barrier to the most basic form of safety evaluation. A pharmaceutical company is required to disclose the chemical composition of its products. An AI educational platform should be required to disclose the architecture, training data provenance, fine-tuning objectives, and behavioral guardrails of its AI components. Without this, efficacy evaluation is possible but safety evaluation is not — and in education, as in medicine, safety is logically prior to efficacy.

The Equity Auditing Gap

An intervention that produces average gains may simultaneously be widening gaps for learners from underrepresented groups, English language learners, students with disabilities, or students from low-income households. Average treatment effects are the wrong primary metric for educational AI that claims to personalize — because personalization implies differential treatment, and differential treatment requires differential evaluation. The WWC reports subgroup analyses where data permits, but it does not require them. An AI-era clearinghouse must require equity auditing as a precondition for any approval, not as an optional supplementary analysis.

1. Versioning and temporal stability. Evidence must be tied to a specific, immutable version of the AI system. Any substantive model update must trigger re-review, as a generic drug reformulation would trigger FDA re-review.

2. Learner model validity. The psychometric soundness of the system's inference about learner state must be independently evaluated — not just whether outcomes improved, but whether the adaptive mechanism was theoretically and empirically defensible.

3. Algorithmic transparency disclosure. Vendors must disclose architecture, training provenance, fine-tuning objectives, and behavioral constraints as a precondition for clearinghouse review — not as a proprietary concession, but as a public accountability requirement.

4. Mandatory equity auditing. Differential impact across learner subgroups must be a required primary analysis, not an optional supplementary one. Approval should be conditional on demonstrated equitable efficacy or an explicit scope limitation.

What the Institution Must Be

The institution we need is not simply a better-funded WWC. It requires a hybrid architecture that draws on the FDA's pre-market authorization logic, the WWC's evidence synthesis and practitioner communication infrastructure, and new capabilities that neither possesses — capabilities specifically designed for the properties of AI systems operating on developing human minds.

Phased Authorization: Borrowing from the FDA

The FDA's phased trial structure — Phase I (safety), Phase II (preliminary efficacy in small populations), Phase III (efficacy in large, representative populations), Phase IV (post-market surveillance) — is not merely procedural bureaucracy. It embeds a logic: that the burden of proof escalates with the scale of deployment and the severity of potential harm. An AI tutoring system deployed to ten thousand students in a controlled pilot carries different stakes than the same system deployed to ten million students across a national curriculum. The authorization framework should reflect this. **Scale of deployment should be gated by quality of evidence.** Broad deployment should require Phase III-equivalent evidence. Pilot use in controlled research contexts should require only Phase I-equivalent safety review.

Evidence Synthesis and Practitioner Communication: Borrowing from the WWC

The WWC's most important contribution is not its evidence standards — it is its translation function. It takes technically complex research findings and renders them legible to practitioners who do not have the methodological training to evaluate primary studies. An AI-era clearinghouse must perform the same translation function, but for a more complex evidence landscape. Educators need to know not just "does this system produce learning gains?" but "for which learners, under which conditions of educator integration, at what cost to self-regulation and metacognitive development, and with what equity implications?" This is a substantially more demanding communication task, and it requires the institutional commitment and resources to execute it at scale.

Novel Capabilities Required

A technical auditing division capable of reviewing AI system architecture, training data, and behavioral constraints — analogous to the FDA's chemistry and manufacturing review capacity. This requires computer scientists, AI safety researchers, and psychometricians working in institutional partnership, not as external consultants.

Post-deployment surveillance infrastructure — xAPI/LRS-compliant data collection standards that enable ongoing monitoring of system behavior and learning outcomes at population scale, without requiring vendors to self-report. As in pharmacovigilance, the institution must have independent access to deployment data, not merely vendor-supplied summaries.

A recall and de-authorization mechanism — the authority to remove a product from approved status when post-deployment evidence reveals harm, equity failure, or behavioral drift that invalidates prior evidence. This authority does not currently exist in educational technology governance anywhere in the world. It must.

FUNCTION	FDA ANALOG	WWC ANALOG	NEW REQUIREMENT FOR AI-EDTECH
Pre-deployment review	Pre-market approval, phased trials	Evidence review before recommendation	Phased authorization gated by deployment scale
Evidence threshold	RCT for efficacy; safety data for approval	RCT/QED tiered evidence standards	+ Learner model validity; equity auditing required
Transparency requirement	Full compositional disclosure	Study design transparency	AI architecture, training provenance, behavioral guardrails
Post-market surveillance	Phase IV; adverse event reporting	No ongoing surveillance mechanism	xAPI/LRS-based ongoing monitoring; versioning tracking
Recall authority	Yes — FDA can withdraw market approval	No — only advisory	De-authorization authority required
Practitioner communication	Prescribing information; black-box warnings	Evidence snapshots; practice guides	Condition-specific, subgroup-specific, educator-role-specific guidance

Completing the Argument

The three essays in this collection form a single argument. The first established that AI in education is constitutively different from AI in any other domain — because its subject matter is human intelligence itself. The second established that educators must therefore be sophisticated consumers of AI, not its developers — as physicians are consumers of pharmaceutical knowledge, not its producers. This third essay establishes the corollary that the second argument cannot stand without: **the physician can only be a responsible consumer if the FDA exists.**

Without an authoritative, independent clearinghouse that evaluates AI-enabled educational tools against rigorous, AI-appropriate evidence standards — before deployment at scale, and continuously after — the educator-as-consumer model is not a safeguard. It is a fiction. We would be asking educators to make evidence-based decisions about AI tools for which no credible evidence base has been systematically constructed and made ac-

cessible. We would be asking them to do, with inadequate tools, what we have just argued they should not have to do themselves.

The institutional gap is not a technical problem. It is a political and normative one. The pharmaceutical industry resisted the FDA. The edtech industry will resist its equivalent. The resistance will be framed in the language of innovation, agility, and the cost of regulatory friction. These are not trivial concerns. Regulatory overhead is real. Innovation timelines matter. But the relevant comparison is not between a regulated and an unregulated market. The relevant comparison is between a regulated market in which educators can responsibly prescribe evidence-validated tools, and an unregulated one in which millions of students are experimental subjects in an uncontrolled trial — without their knowledge, without their consent, and without any mechanism for detecting or correcting the harm.

CONCLUSION / 结语

The physician analogy is only as strong as the FDA that backs it. An educator who is a sophisticated consumer of AI-enabled educational tools is only as responsible as the clearinghouse that has done the work she cannot do alone — that has evaluated the tool's learning theory, validated its psychometric foundations, audited its equity impact, tracked its behavioral stability over time, and communicated its conditions of appropriate use in language she can act on.

We need that institution. It does not yet exist. Building it is not an optional supplement to the reform agenda in educational AI. **It is the precondition for the entire reform agenda.** Without it, every other argument about responsible AI in education — about educator agency, about learner protection, about the constitutive role of human intelligence in learning — remains aspirational rhetoric rather than institutional reality.

The FDA did not emerge from the pharmaceutical industry's voluntary commitment to patient safety. It emerged from documented harm, political will, and the recognition that some risks are too consequential to be governed by market forces and individual professional judgment alone. The history of AI in education will, in the absence of intervention, follow the same path. The question is only whether we wait for the documented harm before acting — or whether, this time, we build the institution before the harm accumulates.