

AI in Education Is Constitutively Different

教育人工智能的根本性差异

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

为何教育AI不能超越人类智慧而独立存在，以及教育者为何应是AI的明智使用者而非开发者。

Xiangen Hu · 胡祥恩

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.

Xianguan 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 II · CHINESE VERSION · 第二部分 · 中文版

教育人工智能的根本性差异

中文立场论文

根本性立场

教育人工智能的根本性差异

为何教育AI不能超越人类智慧而独立存在，以及教育者为何应是AI的明智使用者而非开发者。

胡祥恩 · 香港理工大学应用社会科学系 · 教育研究中心主任

学然后知不足，教然后知困。知不足，然后能自反也；知困，然后能自强也。故曰：教学相长也。

——《礼记·学记》

教，是一切文明赖以延续的行为。它不是信息的搬运，不是内容的灌注，而是一个人的智识在与另一个人的智识相遇时所发生的转化。这个转化，不可能在无人的空间中自动发生。任何试图绕过这一本质的技术，无论多么精妙，都不是教育，只是教育的幻象。

第一论 · 本体论

教育AI的构成性本质：不能超越人类智能而独立存在

在人工智能已证明具有变革性力量的领域——蛋白质结构预测、金融套利、自动驾驶、影像诊断——AI系统作用于明确定义的输入/输出结构。领域对象独立于使用者的心智。一个折叠的蛋白质，无论是否有人理解它，它都在那里。**这些领域的对象，相对于人类认知而言是惰性的**：它们的存在，不需要一个心灵去学习它。

教育是唯一一个**以人类智能的生长本身为对象**的领域。这不是一种设计选择，而是这一事业的定义。没有学习者，就没有学习；没有心灵的转化，就没有教学。将人类智能从教育中抽空，你所剩下的不是退化的教育经验，而是根本没有教育经验。你只剩下信息传输，正如一排书架不是教师一样。

在其他每一个领域，AI可以在执行时无需人类智能在场而独立产生价值。在教育中，人类智能不是「在环路中」——它**本身就是**环路。移去它，环路就不复存在。

构成性区分 · THE CONSTITUTIVE DISTINCTION

这一区分有其精确的理论根基。维果茨基的**最近发展区**确立了：学习只发生在学习者独立能力的边界与在更有知识的他者协助下所能达到的能力之间——这是一个不可还原的双元关系性结构。杜威的经验交互理论坚持认

为，教育不是从一个容器向另一个容器传递内容，而是通过共同探究共同建构意义。克拉克与查尔默斯的延展心灵假说，应用于教育时，意味着AI可以成为真正的认知支架——但**仅仅作为既有人类智能的延伸**，而非其替代品。

独学而无友，则孤陋而寡闻

——《礼记·学记》学习本质上是一种关系性行为。AI可以成为这种关系的延伸，但不能取代这种关系本身。

HAA 框架 - 分类体系

在人类能动性增强（HAA）框架中，AI部署可被分为三类：**甲类**（替代人类能动性）、**乙类**（在离散任务中辅助人类能动性）、**丙类**（以增强认知触达和自主性的方式延伸人类能动性）。在大多数领域中，三种类型在适当语境下均可合法存在。

在教育中，**甲类部署不仅是次优选择——而是一种范畴性错误**。只有丙类——AI作为教育者-学生关系的延伸——才是本体论上适当的。这不是政策偏好，而是对教育本质的逻辑推论。

比较维度	其他领域的AI	教育中的AI
领域对象	独立于人类认知（蛋白质、价格、像素）	人类认知本身——学习者正在发展中的心灵
AI能否独立运作？	✓ 能 — 无需人类在场亦可产生价值	✗ 不能 — 没有教育者/学生，就没有教育行为
与人类智能的关系	可选的增强或替代	构成性要求 — AI本身就是延伸
失败模式	次优输出，可纠正	范畴性失败 — 教育关系本身瓦解
适当的HAA类型	甲、乙或丙，视语境而定	仅丙类 — 延伸，永不替代
首要问责主体	工程师、领域专家	教育者 — 对学习者的认识论责任

第二论 · 专业论

专业主张：教育者应为明智的使用者，而非开发者

当下教育技术文化中，已悄然滋生出第二种错误：认为教育者不仅应当采用AI工具，更应当**亲自构建**它们。低代码、零代码AI平台的兴起，大语言模型的普及化，以及教育共同体渴望创新的热情，共同制造了一种危险的幻觉——好像教学法专业知识足以胜任教育AI的开发工作。事实并非如此。医学的历史为我们提供了一个澄清性的类比。

医者之道：知药而不制药

医生不合成药物，不设计临床试验，不建立生物利用度。这些工作属于药理学家、化学家、生物统计学家。医生所做的是另一件更难、更重要的事：**将那些证据应用于站在眼前这个不可还原的特殊患者**。她知道何时开药、何时不开药、何时需要转诊。她的专业知识是**判断**，而非合成。

术业有专攻

——韩愈《师说》 各有专门之业，各守其职，各尽其责。教育者的专攻，是教育判断，而非AI工程。

医学领域

医生的角色

- 理解药理学证据
- 评估随机对照试验与系统综述
- 为患者做出处方决策
- 不合成药物
- 不自行开展临床试验
- 对患者结局承担认识论责任

教育领域

教育者的角色

- 理解学习科学证据
- 评估效果研究与IRT心理测量效度
- 为学习者做出教学决策
- 开发AI系统
- 自行开展效度验证
- 对学习者发展承担认识论责任

如果医学院鼓励医生自行合成药物，以「医生最了解患者」为由，在没有临床试验的情况下直接用于患者，我们会称之为危险。对其教育领域的等价行为，我们理应给出相同的判断。

专业类比 · THE PROFESSIONAL ANALOGY

教育者自主开发AI的认识论风险

热情偏见与新奇效应。教育工作者同时扮演开发者、使用者和评估者三种角色——这种三重利益冲突使客观的效果评估几乎成为不可能。

心理测量基础的缺失。有效的教育AI需要根植于项目反应理论（IRT）、学习者特征曲线（LCC）、知识空间理论等框架。没有心理测量基础的教育者构建自适应工具，实际上只是在**模拟**自适应，同时系统性地偏向已符合其直觉模型的学习者。

伦理与发展效度的缺口。与学习者互动的AI系统，尤其是儿童与青少年，承载的风险不只是技术性的，更是发展性的和伦理性的。认知负荷、动机架构、长期元认知发展等问题需要专业知识才能预见。

不可复制性问题。教育研究的证据标准——预注册、控制条件、效应量估计——需要制度性基础设施与方法论训练。在这一基础设施之外产生的工具，充其量产生的是轶事，而非证据。

走向新标准：先有效果证据，再谈规模推广

制药行业运行在一个不可逆转的制度规范之下：**没有任何药物能在没有结构化证据门槛的情况下到达患者手中**。然而没有任何等价的规范约束教育AI。工具以软件产业的速度被发布、营销、采用和规模化推广，却几乎不承担软件业对迭代失败所持有的容忍度的对应责任。

软件漏洞让程序崩溃。教育AI的漏洞塑造认知、扭曲自我效能感，并可能制造持久的学习脱离或根深蒂固的错误概念。部署速度与发展影响之间的不对称，是一种文明层面的风险。

我们所需要的标准包括：**理论基础在先**（每一个教育AI系统都必须阐明其学习理论）；**心理测量效度是前提**（自适应系统必须证明其学习者模型的效度）；**效果证据先于规模化**（带有适当对照条件的先导研究，尽可能预注册）；以及**部署后的持续监测**（正如药物警戒体系，持续监测非预期效应）。

CONCLUSION / 结语

教育人工智能在本质上不同于其他任何领域的人工智能。这种差异源于关于教育最基本的事实：**教育的对象，是人类心灵在成为其更完整自身的过程中**。

正因如此，教育中的AI不能超越人类智能而独立存在。它没有任何可以作用的对象，除了正在发展的心灵；而正在发展的心灵，需要一位人类教育者——有意图、有知识、有伦理担当——始终处于学习关系的中心。AI延伸了那种关系，它不能替代那种关系。

正因如此，教育者不是AI开发者。他们是，或者必须成为，AI的老练而要求严格的使用者——能够评估证据，识别危险信号，并将AI系统置于我们对其他任何作用于人类健康与发展的干预措施所要求的同等效果验证标准之下。

医生不制造药物，但她知道何时开药、何时不开药、何时它正在造成伤害。这是那个模型，这是那个标准。**教育现在必须有勇气接受它**。

框架参考 · 人类能动性增强框架 (HAA) · 甲/乙/丙类分类 · 最近发展区 (维果茨基, 1978) · 延展心灵 (克拉克与查尔默斯, 1998) · IRT基础 (Lord, 1980) · 循证教育 (Slavin, 2002) · 学习者特征曲线 (LCC) · AutoTutor (Graesser & Hu)