

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

两个论点——教育中的AI不能超越人类智能而独立存在；以及针对AI教育资源建立「药监局」式证据审查机构是一项逻辑必然——作为一份完整立场论文呈现。

目 录

第一篇 · 教育人工智能的根本性差异

第二篇 · 医者类比：没有「药监局」便不完整

胡祥恩

香港理工大学应用社会科学系讲座教授

教育研究中心主任

人类能动性增强框架（HAA）· 2026年4月



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)



第四部分 · 中文版 · PART IV · CHINESE VERSION

医者类比：没有「药监局」，类比便不完整

关于建立AI教育资源证据审查机构的制度性提案

制度性推论

医者类比：没有「药监局」便不完整

「教育者应是AI的使用者而非开发者」这一论点，只有在严格的验证制度存在的前提下才能成立。我们需要一个针对AI教育资源的「药监局」与「有效证据数据库」——而现有机构均不足以胜任。

胡祥恩 · 香港理工大学应用社会科学系 · 教育研究中心主任

名不正则言不顺，言不顺则事不成，事不成则礼乐不兴。

——《论语·子路》

前两篇文章确立了两个主张：教育中的AI不能超越人类智能而独立存在；因此，教育者必须是AI的明智使用者，而非其开发者。两个主张皆为真。但合而观之，它们产生了一项尚未被兑现的义务。

医者类比，细察之下，隐含一个前提。我们能够要求医生做一个负责任的处方者——评估证据、运用判断、保护患者——**当且仅当药监局存在**。医生作为药学知识消费者的能力，预设了一套制度基础设施已经完成了艰难的工作：在受控条件下评估了安全性与有效性，建立了证据门槛，具备撤回有害产品的权威，并将其发现以从业者可读的形式公开。没有药监局，要求医生「只需做个好的使用者」不是一种保障，而是一种卸责。

我们不能告诉教育者要负责任地使用AI，除非我们同时建立使负责任的使用成为可能的制度基础设施。医者类比要求药监局的存在；教育者类比要求其等价机构的存在。

制度性推论 · THE INSTITUTIONAL COROLLARY

教育领域对AI教育工具没有这样的基础设施。美国教育科学研究所（IES）运营的「有效教育干预数据库」

（What Works Clearinghouse, WWC）是最接近的现有类比——它有价值、有严格性、有必要性。但它是为前AI时代的证据格局设计的。它评估结构化课程和离散干预，使用随机对照试验和准实验设计。它不是为评估以下对象而设计的：根据学习者状态改变行为的自适应系统；输出具有随机性和语境依赖性的大语言模型；或底层模型可能在部署后被悄然更新的AI平台。WWC所做之事与AI时代的审查机构所必须做之事之间的差距，不是程度之差，而是种类之差。

WWC没有被设计来处理的问题

WWC的证据标准是教育研究中最严格的之一。其分层框架——强证据、中等证据、初步证据、无证据——比任何其他单一制度机制都更有效地约束了教育决策。它理应享有其权威。此处的论点不是反对WWC，而是AI教育工具呈现出一套WWC当前方法论无法处理的截然不同的验证挑战。

稳定性问题

传统的教育干预是稳定的：2023年实施的阅读课程，与2024年实施的是同一课程。其属性可以被测量，测量结果随时间保持有效。由大语言模型支撑的AI教育工具则不然。模型可以被供应商更新、微调或替换——悄然进行，无需通知学校或监管机构——以可能使任何先前效果证据失效的方式改变系统行为。一种2023年获批的药物不能在2024年自发改变其分子结构。一个AI辅导系统可以，且经常这样做。**对一个「活系统」的效果证据，需要一套教育研究从未需要过的版本控制与监测框架。**

自适应效度问题

自我标榜为「自适应」的教育AI系统正在迅速增多。几乎每个主要的教育技术平台都声称能够根据学习者状态个性化教学。但自适应只有在底层学习者模型有效的情况下才具有教育价值——即系统关于学习者所知、所能及下一步需要什么推断，必须建立在可辩护的心理测量理论之上，而非行为代理指标、参与度指标或点击流模式。WWC的当前框架没有评估学习者模型效度的机制。它能评估使用了某个系统的学生是否优于未使用的学生，但无法评估系统的自适应逻辑是否是这一结果的原因。

黑箱透明度问题

许多AI教育系统——尤其是建立在专有大语言模型上的——以黑箱方式运作。其推荐逻辑、内容生成参数和反馈机制对教育者、研究人员或监管机构均不可检视。这不仅仅是研究者的不便，而是对最基本形式的安全评估的认识论障碍。制药公司须披露其产品的化学成分；AI教育平台应当被要求披露其AI组件的架构、训练数据来源、微调目标和行为约束。没有这些，效果评估是可能的，但安全评估则不然——而在教育中，一如在医学中，安全性在逻辑上先于有效性。

公平审计缺口

一项产生平均收益的干预，可能同时正在扩大来自弱势群体学习者、英语学习者、残障学生或低收入家庭学生的差距。平均处理效应对于声称要个性化的教育AI而言是错误的主要指标——因为个性化意味着差异化处理，而差异化处理需要差异化评估。针对AI教育工具的审查机构必须将公平审计作为任何批准的前提条件，而非可选的补充分析。

1. **版本控制与时间稳定性。**证据必须与AI系统的特定、不可变版本挂钩。任何实质性的模型更新都必须触发重新审查，就像仿制药重新配方需要触发药监局重新审查一样。
2. **学习者模型效度。**系统关于学习者状态的推断的心理测量合理性必须被独立评估——不仅仅是结果是否改善，而是自适应机制在理论和实证上是否站得住脚。
3. **算法透明度披露。**供应商必须将架构、训练来源、微调目标和行为约束作为审查的前提条件——不是专有的让步，而是公共问责要求。
4. **强制性公平审计。**对不同学习者子群体的差异化影响必须是必需的主要分析，而非可选的补充。批准应以证明公平有效或明确范围限制为条件。

这个机构必须是什么

我们需要的机构不是简单地获得更多资金的WWC，而是需要一种混合架构：借鉴药监局的上市前授权逻辑，借鉴WWC的证据综合与从业者沟通基础设施，以及两者都不具备的新能力——专门为在发展中的人类心灵上运作的AI系统的特性而设计。

分阶段授权：借鉴药监局

药局的分阶段试验结构——I期（安全性）、II期（小样本初步有效性）、III期（大规模代表性样本有效性）、IV期（上市后监测）——嵌入了一个逻辑：举证责任随部署规模和潜在伤害严重性的增加而提升。**部署规模应以证据质量为门控条件。**广泛部署应要求III期等效证据；在受控研究语境中的试点使用，只需I期等效安全审查。

证据综合与从业者沟通：借鉴WWC

WWC最重要的贡献不是其证据标准，而是其翻译功能：将技术上复杂的研究发现转化为没有方法论训练的从业者可以理解的语言。AI时代的审查机构必须执行同样的翻译功能，但面对更复杂的证据格局。教育者需要了解的不仅仅是「这个系统是否产生学习收益」，还要知道：对哪些学习者、在哪些教育者整合条件下、以对自我调节和元认知发展何种代价、以及有何公平含义。这是一项要求更高的沟通任务。

所需的新能力

技术审计部门——能够审查AI系统架构、训练数据和行为约束，类似于药监局的化学与制造审查能力。这需要计算机科学家、AI安全研究人员和心理测量学家以制度性伙伴关系工作，而非作为外部顾问。

部署后监测基础设施——符合xAPI/LRS标准的数据收集规范，能够在不要求供应商自我报告的情况下，对系统行为和学习结果进行持续的人口规模监测。如同药物警戒，机构必须独立获取部署数据，而非仅依赖供应商提供的摘要。

撤销与取消授权机制——当部署后证据揭示伤害、公平失败或使先前证据失效的行为漂移时，将产品从已批准状态中移除的权威。这种权威目前在世界任何地方的教育技术治理中都不存在。它必须存在。

功能	药监局类比	WWC类比	AI教育技术新要求
部署前审查	上市前批准，分阶段试验	推荐前证据审查	以部署规模为门控的分阶段授权
证据门槛	有效性RCT；安全数据	RCT/准实验分层标准	+学习者模型效度；强制公平审计
透明度要求	完整成分披露	研究设计透明度	AI架构、训练来源、行为约束
上市后监测	IV期；不良事件报告	无持续监测机制	基于xAPI/LRS的持续监测；版本追踪
撤销权威	有 — 药监局可撤销市场批准	无 — 仅为咨询性	取消授权权威为必要条件
从业者沟通	处方信息；黑框警告	证据快照；实践指南	针对条件、子群体、教育者角色的专项指导

完成论证的闭环

格物、致知、诚意、正心、修身、齐家、治国、平天下。

——《大学》

本系列三篇文章构成一个完整的论证。第一篇确立：教育中的AI在本质上不同于其他任何领域的AI——因为其对象是人类智能本身。第二篇确立：教育者因此必须是AI的明智使用者，而非其开发者——如同医生是药学知识的使用者，而非其生产者。第三篇——本篇——确立了第二个论点不可缺少的推论：**医生只有在药监局存在的情况下才能成为负责任的处方者。**

如果没有一个权威的、独立的审查机构，以严格的、AI适用的证据标准——在大规模部署之前，以及部署之后持续地——对AI教育工具进行评估，那么「教育者作为使用者」的模式就不是一种保障，而是一种虚构。我们将要求教育者对没有可信证据基础的AI工具做出循证决策——我们将要求他们用不足的工具，去做我们刚刚论证他们不应该自己做的事情。

制度空白不是一个技术问题，而是一个政治与规范问题。制药行业曾经抵制药监局；教育技术行业将抵制其等价机构。抵制将以创新、敏捷性和监管摩擦成本的语言被框架化。这些不是无关紧要的顾虑。监管开销是真实的，创新时间线很重要。但相关的比较，不是受监管市场与不受监管市场之间的比较。相关的比较，是一个教

育者可以负责任地开出经过证据验证的工具处方的受监管市场，与一个数百万学生在一项不受控的试验中成为实验对象的不受监管市场之间的比较——在他们不知情的情况下，在没有知情同意的情况下，在没有任何检测或纠正伤害的机制的情况下。

CONCLUSION / 结语

医者类比只有在支撑它的药监局存在的情况下才足够有力。一个负责任地使用AI教育工具的教育者，只有在一个已经完成了她独自无法完成的工作的审查机构的支撑下，才能真正负责任——这个机构评估了工具的学习理论，验证了其心理测量基础，审计了其公平影响，追踪了其随时间的行为稳定性，并以她能够采取行动的语言传达了其适当使用条件。

我们需要那个机构。它尚不存在。建立它不是教育AI改革议程中的可选补充。**它是整个改革议程的前提条件。**没有它，每一个关于教育中负责任的AI的论点——关于教育者能动性、关于学习者保护、关于人类智能在学习中的构成性角色——都只是停留在志向的修辞，而非制度的现实。

药监局不是从制药行业自愿承诺患者安全中产生的。它是从有记录的伤害、政治意志，以及对某些风险太过重大、不能仅由市场力量和个人专业判断来治理这一认识中产生的。在缺乏干预的情况下，教育中AI的历史将沿循同样的路径。唯一的问题是：我们是等待伤害被记录后再行动——**还是这一次，在伤害积累之前，先建立那个机构。**