

作业之死：教育评价对象的蒸发与发生过程的重新显影

学校正在用越来越精密的手段防止学生用机器完成作业。这场战争的全部前提是：作业仍然是学习发生过的证据。这个前提已经不成立了——而战争还在打。

王德生 著 · 学术创造专栏 · 典范论文 · 约 8000 字

摘要

生成式人工智能进入课堂之后，教育界的第一反应是把它定性为一场作弊危机，并投入大量资源建设检测、留痕与监考装置。本文论证这个定性是错的，而且错在根部。作业从来不是学习的目的，它是学习这场发生留下的显露态——教师从未直接看见学习，他看见的是作业，并由作业反推那条看不见的差异序列。机器没有伪造这个证据，它做了更彻底的事：它让显露态与它本该显露的那场发生第一次相互脱钩。学生交上来的东西可以完美无缺，而它本该指证的路径根本没有走过。于是评价面对的不是欺骗，是对象的蒸发——被评价的那个东西已经不在，评价还在照常进行，并且照常给出分数。本文进一步论证，"过程留痕"（记录编辑历史、要求提交草稿、锁定考场）不能修复这一点，因为它仍然在寻找一个物证，而蒸发的恰恰是物证与事实之间的那条联系。真正的出路是把评价对象从结果表征迁移到发生过程本身的显影，并指出这一迁移需要一种此前不存在的装置：一种不替学生思考、而让学生的差异序列变得可见、可回放、可判错的复合智能体。文末给出该方案自身的四个失效条件。

关键词 教育评价；生成式人工智能；认知外包；发生学；过程显影；复合智能体；判错权

二〇二五年秋天，一位大学教师在批一份课程论文时遇到了一个新问题。这份论文挑不出毛病：论证清晰，文献准确，结构完整，甚至有一段关于研究局限的自我反思写得相当漂亮。查重率百分之二。检测工具给出"人类撰写概率百分之七十八"——一个既不能定罪也不能免罪的数字。他最后打了八十八分，然后在办公室里坐了很久。

后来他说，让他不安的不是"这可能是机器写的"。让他不安的是，他发现自己回答不了一个更靠前的问题：

如果这确实是学生自己写的，我凭什么知道？

这个问题在二〇二二年以前不需要回答。不是因为那时候有答案，而是因为那时候没有人会这样问。一份写得好的作业和一场发生过的学习之间的联系，在过去两千年里牢固到没有人觉得它是一个需要论证的命题。它就是教育本身的地基。

现在这块地基裂了。裂的方式很特别：它不是被砸开的，是被抽空的。上面的建筑一点没变，会议照开，作业照收，分数照给，检测软件的市场每年增长四成——只是承重的那样东西不在了。

一、这不是一场作弊战争

把生成式人工智能的冲击定性为作弊，是教育界迄今为止最省力、也最昂贵的一次判断。省力，是因为作弊是一个已有的范畴，学校处理它有两千年的经验：抄袭、代考、夹带、买卖论文——每一种都有对应的规则、检测手段与惩戒程序。把新东西塞进旧范畴，整个系统一天都不用停就能开始应对。昂贵，是因为这个范畴恰好错在最关键的地方。

作弊有一个严格的结构：**存在一个真实的应然过程，行为人用一个伪造的表征去冒充它**。代考的前提是“本人考试”这件事是可能的且是标准；抄袭的前提是被抄的那个原创过程真实发生在别人身上。作弊是关于**表征与事实之间的欺骗关系**——事实在那儿，只是被冒名顶替了。

正因如此，反作弊在逻辑上永远是可行的：既然事实存在，就总能找到把表征与事实重新对上的办法。摄像头、笔迹、时间戳、指纹、现场重述——反作弊的历史就是不断加固表征与事实之间那条联系的历史。

生成式人工智能不在这个结构里。学生用它写论文的时候，没有一个“本该由他完成的真实过程”被冒名顶替。**那个过程根本没有发生**。这里没有事实被欺骗，这里是事实缺席，而表征照常产出。

差别听起来细微，后果是决定性的：

作弊是表征与事实的错配，可以通过加固联系来解决；这件事是事实的缺席，加固联系只会让你更快地确认那头是空的。

这解释了一个让所有人困惑的经验事实：检测技术越精密，教育者越不安。如果这真是一场作弊战争，技术进步应该带来安心才对。它没有。因为技术每精确一分，都在把同一个空洞照得更亮一点。

于是要问的不是“怎么抓住他们”。要问的是：**我们本来在评价什么？**

二、作业曾经是一根指针

回答这个问题，得先承认一件教育界很少说破的事：**教师从来没有直接看见过学习**。

学习是发生在一个人的纠缠里的事——他已有的经验、他此刻的困惑、他昨天读到的一句话、他和同伴争论时突然卡住的那半秒。这些东西不可观测，不可传递，也不可复制。教师能看见的只有一样：**这场发生留下的显露态**。作业、答卷、发言、眼神、犹豫的长度。

作业的全部价值就在这里。它不是学习的目的，它是学习的**指针**——一个可以被看见的东西，指向一个不可被看见的东西。教师批改作业时干的活，本质上是一次反推：从显露态反推那条看不见的差异序列。这个学生在第三步卡住了，说明他没建立起那个区分；那句表述别扭，说明他是靠着模仿而不是靠着理解写出来的；这个反例举得刁钻，说明他真的在拿这个概念去撞现实。

好教师的所谓“经验”，说穿了就是这套反推的精度。他能从一份作业里读出的东西，远远超过作业上写着的东西。

这套反推之所以可靠，靠的是一个隐含的物理约束：**在过去，要产出这个显露态，就必须走过那条路径。**没有捷径。你要写出一篇讲得通的分析，就必须真的想过；你要解出这道题，就必须真的建立起那些区分。**显露态是路径的函数——这不是一条教育规律，这是一个当时的事实。**正因为它是事实，指针才可靠，教育评价才成立。

这个事实持续了两千年，从来没有被写进任何一本教育学教科书。它太理所当然了，理所当然到没有人觉得它是一个需要被命名的东西。

三、指针与被指的东西第一次分离

现在这个事实没了。

生成式人工智能干的事，用最精确的话说，**是把显露态从它的发生路径上解耦了。**它可以在完全不经过那条路径的情况下，产出那条路径本该产出的东西——而且往往产得更好，因为它见过一百万条相似的路径的结果。

请注意，它并不是"模仿"了那条路径。它没有走一条更快的路，它走的是另一条完全不同的路：一条在符号层面上重组沉淀物的路。这条路走完之后产出的显露态，与走过原路径产出的显露态，在文本上无法区分——不是"很难区分"，**是没有区分可言**，因为两者本来就是同一类符号排列。

于是那根指针断了。它还在，还在指，但它指向的那一头是空的。

显露态从此不再是路径的证据。它只是显露态。

这句话是这篇文章的整个重量所在。让我把它的后果一条条摊开，因为教育界的讨论至今没有一条走到这里：

其一，全部现存的教育评价工具，无一例外，评的都是显露态。从考试到论文，从作品集到答辩，从"表现性评价"到"真实性评价"——这些花样翻新的改革方案争论的只是"看哪一种显露态"，从来没有争论过"显露态还指不指得住"。它们全部预设那根指针是好的。

其二，"更真实的任务"救不了它。教育界最流行的应对是"设计 AI 做不了的任务"：让学生做本地调研、做小组辩论、写个人反思。这条路的寿命由模型能力单方面决定，而且它已经在缩短。更要命的是它的逻辑还在原地：它仍然在找一种"机器走不通、人走得通"的显露态，也就是仍然假设指针可以被修好。它没有承认指针的断裂是**结构性的**——只要显露态可以被另一条路径产出，这根指针在原理上就不可靠，剩下的只是时间。

其三，也是最难接受的一条：这不是学生的道德问题。一个从不用 AI 的学生和一个全程用 AI 的学生，交上来的东西可以完全一样。评价系统对二者给出同一个分数，而且这个分数是"公正"的——因为评价系统看的就是这个东西，而这个东西确实一模一样。系统没有出错，系统在完美地执行它的规则。**出错的是规则赖以成立的那个世界。**

四、蒸发：不是证据被伪造，是对象不在了

到这里可以给出这篇文章的核心判断了。

学校现在面对的，不是一个证据被伪造的问题，而是**评价对象的蒸发**：

评价制度声称评价的那个东西——学习的发生——已经不在它所能触及的范围内了；而评价还在照常进行，照常给出分数，照常发放学位，并且照常自认公正。

"蒸发"这个词是精确的，不是修辞。蒸发的特征是：没有爆炸，没有残骸，容器完好无损，只是里面的东西不见了。你把手伸进去，什么都摸不到，但从外面看，容器还是那个容器，摆在原来的位置上，标签也没换。

一个蒸发了对象的评价系统会怎么运行？它会继续运行，而且运行得比以前更顺畅。这不是悖论，这是必然：

分数会更高。机器产出的显露态在质量上高于平均学生。评分标准是对着显露态定的，所以分数普涨。教务系统看到的是"教学质量提升"。

方差会更小。学生之间的差别主要来自路径，路径外包之后，显露态之间的差别会向工具能力收敛。教务系统看到的是"教学质量趋于均衡"。

投诉会更少。评分争议来源于教师的反推是有偏差的；不再反推、只看显露态之后，评分变得更"客观"。教务系统看到的是"评价公信力上升"。

三项指标同时向好，而被评价的那个东西已经不在了。**这才是最深的一层：蒸发不但不制造警报，它制造捷报。**任何一个只靠指标运行的系统，都会把这场蒸发读成一次成功。

这就是为什么这件事不能靠"加强管理"来处理。管理是靠指标运行的，而这场事故的指标全是绿的。

五、"过程留痕"是同一个错误的更贵版本

目前呼声最高的方案是"过程留痕"：用能记录编辑历史的写作平台，要求提交草稿、大纲、参考文献笔记，用版本回放证明这篇东西是一点点长出来的。有的学校已经开始强制安装键盘记录插件。

这个方案的直觉是对的：它意识到了成品不再可信，所以想去看过程。它的执行是错的，而且错在同一个位置。

因为它要的仍然是一个**物证**。它只是把物证从"成品"换成了"成品的成长史"。而成长史同样是一个显露态，同样可以被另一条路径产出——事实上，模拟一份合理的编辑历史比写那篇论文本身更容易，市面上已经有工具在做这件事，而且它做得很好，因为"渐进式编辑"是一个比"好文章"简单得多的模式。

留痕方案的真正代价还不在这里。它的代价是**它把师生关系重新定义成了一次取证**：

一个被键盘记录器监视着写论文的学生，正在经历的**不是学习，是审讯**。而学习需要的东西里，恰好有一样在审讯中无法存在：**允许自己写出蠢话的余地**。思考的路径本来就是由大量的错、乱、绕远、推倒

重来构成的——这些正是介生态的样子：目标半成形，路径半试探，什么都还没定。留痕把这段过程变成了呈堂证供，于是每个学生学会的第一件事就是：先在别处想好，再到留痕系统里表演一遍思考。

这不是留痕的副作用，这是留痕的必然产物。你把过程变成证据，过程就会开始表演；表演出来的过程和真实的过程，在留痕系统里同样无法区分——指针又断了一次，只是这次更贵。

凡是把过程当物证来收集的方案，都会在收集的同时把它要证明的那个东西赶走。

五之二、"这不过是又一次计算器"

到这里，一个反驳会准时出现，而且它是这个题目上最强的一个，必须正面接住，绕过去这篇文章就不算数。

反驳是这样的：每一次新工具出现，教育界都恐慌一次，每一次都被证明是虚惊。苏格拉底说文字会毁掉记忆力。十九世纪的教师说印刷的书会让学生不再用心听讲。上世纪七十年代，全美国的数学教师联名反对计算器进课堂，理由和今天一字不差——"学生会失去心算能力，会不理解算术的本质"。今天没有人认为计算器毁了数学教育。二〇〇〇年，教授们说搜索引擎会让学生不再读书、不再记忆、不再思考。今天没有人怀念卡片目录。

每一次，人类都适应了；每一次，事后看恐慌的人都很蠢。凭什么这次不一样？

这个反驳的力量在于它有一整部历史撑腰。而它的错误也正在这里：它对每一次都用了同一个描述——"新工具引发恐慌，后来适应了"——这个描述太粗，粗到把三件性质完全不同的事看成了同一件。

把它们分开看：

计算器接手的是执行。一个学生用计算器算 847×36 ，他外包掉的是乘法的执行。但"该不该乘""乘完这个数意味着什么""这个结果为什么荒谬"——判断全部留在他身上。计算器给出 30492，如果他的题目本来应该是加法，计算器不会告诉他。这就是为什么计算器最终被教育接纳，而且接纳得心安理得：它拿走的那一段，本来就**不是学习发生的那一段**。数学教育在计算器之后不但没死，反而第一次能把课时投到它真正该投的地方——建模、估算、判断合理性。计算器**不是被容忍的，它是被感谢的**，因为它替教育剥掉了一层不该占那么多课时的壳。

搜索引擎接手的是检索。一个学生用搜索代替卡片目录，他外包掉的是"找到相关文献"。但读、判断可靠性、发现两篇文章互相矛盾、决定信谁——全部留在他身上。搜索甚至加重了判断的负担：卡片目录里的东西都是被图书馆员筛过的，搜索结果不是。这就是为什么信息素养教育在搜索引擎之后反而成了一门正经课程——工具把判断的量放大了，于是判断第一次需要被专门教。

现在这一次接手的是什么？

它接手的是那条路径本身。它不是替学生做了学习之前的事（找资料），也不是替他做了学习之后的事（算数字）——它替他做了学习。

分界线在这里，而且它是可以精确划出来的：

计算器和搜索引擎拿走的，是显露态之外的东西，它们让路径更容易走完，路径还是他走的。这一次拿走的，是路径本身，而只留下了显露态。

这个区分不是文字游戏，它有一个可检验的后果：用计算器的学生，你从他的作业里仍然能反推他的路径——他的建模对不对，他的估算合不合理，他有没有发现结果荒谬。指针还在。用生成式模型的学生，你从他的作业里什么也反推不出来，因为那条路径根本没有走过。指针断了。

所以"人类总会适应的"这句话在这里是对的，但它对得毫无用处。人类当然会适应——问题是适应之后剩下的那个东西还是不是教育。一个把路径全部外包、只保留显露态的教育系统，它会平稳运行，它会有分数，会有毕业生，会有排名，会有满意度调查。它只是不再让任何事情在任何人身上发生。而"适应"这个词最狡猾的地方就在于，它把这种结局也算成一次适应。

历史上那几次恐慌之所以事后看很蠢，恰恰因为那几次工具没有碰路径。这一次它直接落在路径上。用前几次的"事后证明是虚惊"来推断这一次，等于说：前三次感冒都自己好了，所以这次也不用管——而这次不是感冒。

六、显影：把评价对象搬到发生过程上

那条出路在哪里？

先要把两个被混为一谈的东西分开：过程留痕和过程显影。

留痕是事后收集物证：它默认过程已经发生完了，任务是把它残骸捡起来做鉴定。显影是当场让发生变得可见：不是收集证据，是提供一个让那条路径能在其中显露出来的场。

差别在哪里？照相机和显微镜的差别。照相机记录已经发生的，显微镜让本来看不见的当场可见——而且被看见的东西并没有因此被改变，因为显微镜不是审讯。

一场学习的发生里，真正带信息的是什么？不是结论。结论是可以外包的，那正是我们刚刚论证过的。带信息的是三样东西：

一是判错。学生在什么地方觉得不对劲。这个"不对劲"不能外包——机器可以给你一百个候选，但"这个不对"这句话必须由一个在场的人说出来，因为不对劲是一种承受，不是一个判断结果。判错权是全部教育里最后一块不可让渡的东西。

二是卡点。他在哪里停住了，停了多久，绕了几圈才过去。卡点是他的纠缠的地形图——同一道题，两个学生的卡点位置不同，说明他们的底盘不同。这个信息在成品上一点痕迹都不留，而它比成品重要得多。

三是撤销。他推翻了自己的哪个想法，为什么推翻。一个从不推翻自己的学习过程，多半没有发生过学习——它只是搬运。

这三样东西合起来，才是那条差异序列本身。把评价对象从"他产出了什么"迁移到"他在哪里判错、在哪里卡住、推翻了什么"，就是把评价对象从显露态搬回发生过程。

这不是一个比喻。这是一个可操作的迁移，而且它有一个决定性的好处：**这三样东西没法外包**。你可以让机器替你写论文，你没法让机器替你觉得不对劲——如果你让它替你觉得，那么"不对劲"这件事就没有发生，而它的缺席是当场可见的，不需要任何检测。

七、复合智能体：让路径可判错的装置

这个迁移不能靠一份新的评分表来完成。它需要一个此前不存在的装置。

原因很简单：这三样东西——判错、卡点、撤销——在传统课堂里是不可观测的。三十个学生一个老师，老师看不见任何一个人的卡点。这正是为什么教育两千年来只能评显露态：不是它想评，是它只看得见这个。

现在第一次有了另一种可能，条件是那个装置被造成正确的样子。

它必须做的第一件事，是承担可外包的那一半。 让机器去写初稿、去查文献、去生成候选——不是因为妥协，而是因为只有把可外包的东西全部外包掉，剩下的那个核才会露出来。一个学生在没有 AI 的时候，他的判错混在他的组织、他的措辞、他的查资料里，谁也分不出来。现在那些全被拿走了，剩下的赤裸裸就是：面对这五个机器给的候选，你说哪个不对，为什么。

复合智能体的第一个功能不是提高产出，是把不可外包的核逼到台面上。

它必须做的第二件事，是把判错权交给学生并且不许收回。 这是整个设计里最容易做错的一条。绝大多数教育 AI 产品在这里失守：它们给答案、给评语、给"你这里错了"。这样的产品每用一次，都在把判错权从学生手里拿走一点——而判错权正是我们刚刚论证过的那个最后不可让渡的东西。一个替学生判错的教育 AI，在做的事情和替学生写论文的 AI 是同一件，只是更隐蔽，因为它看起来像在教。

正确的做法是反的：机器只出候选，不排序，不推荐，不暗示。学生必须自己说哪个不对。他说错了，机器不纠正，机器给出这个判断在现实中的后果——让现实去纠正他。**判错权留在学生一侧，纠错的代价由现实承担，机器只负责把现实搬到眼前。**

它必须做的第三件事，是把卡点和撤销变成过程本身的一部分，而不是被记录的证据。 差别在于用途：留痕是把这些交给教务系统去审；显影是把这些当场还给学生自己看——你在这里绕了四圈，你三次推翻了同一个想法，你从来没有对第二类候选说过"不对"。这些不是罪证，这是他的地形图。而当地形图被还给他自己的时候，一件教育学上从没发生过的事出现了：**学生第一次能看见自己的差异序列。**

这时候评价变成了什么？它变成了对一场可见的发生的判读。教师不再需要反推——他能直接看见那条路径。而学生也不再是被取证的对象，他是那张地形图的第一读者。**取证和显影的分水岭就在这里：证据是给别人看的，地形图是给自己看的；一个制造表演，一个制造发生。**

八、这套东西什么时候失效

任何把话说尽的方案都是自己做完了自己。所以必须在这里留下四道裂缝——它们是这套东西的失效条件，也是后来的人能钻进去的地方。

其一，判错权可以被表演。学生完全可以学会"该在哪里说不对"——如果说"不对"能得分，那么"说不"就会变成新的显露态，然后新一轮蒸发开始。这套方案的寿命，取决于判错的后果能不能真的由现实来结算，而不是由分数来结算。凡是判错的代价最终落在分数上的系统，都会在两年内长出新的表演。这是最要命的一道缝，我不认为它已经被补上。

其二，它对学科不是均匀的。判错、卡点、撤销这三样，在数学、编程、实验科学里是清晰的；在文学、艺术、哲学里，"不对"往往要在很久以后才知道，甚至永远不知道。这套方案在软的一头会退化成为一种伪精确——把"当场说得为什么不对"当成标准，恰恰会淘汰掉那些需要在黑暗里待很久的学生。

其三，它假设学生已经有底盘。一个连"不对劲"都感觉不到的初学者，把判错权交给他等于把他扔在海里。这套东西可能只在中级以上成立，而初级阶段该怎么办，本文没有回答。

其四，也是最深的一条：显影本身会改变被显影的东西。一个知道自己的卡点正在被显示的人，和一个不知道的人，卡的方式不一样。这不是可以通过设计规避的技术问题，这是反身性对象上的知识生产的宿命——凡对象会自我建模处，封顶只能逼近而不抵达。这套方案能做的最多是把这道缝管理好，做不到消除它。

回到那位坐在办公室里的教师。

他的问题——"如果这确实是学生写的，我凭什么知道"——现在有答案了，但答案不是他想要的那种。答案是：凭那份论文，你永远不可能知道，从今以后都不可能。不是因为学生变狡猾了，也不是因为工具变强了，是因为那根让你能够知道的指针，它所依赖的那个物理事实，已经不存在了。

这件事没有回头路。任何试图把指针修好的方案——更严的检测、更真实的任务、更细的留痕——都是在修一根断掉的指针，而它断的地方不在指针上，在世界上。

但还有另一件事也是真的，而且它比前一件更值得被说出来：在过去两千年里，教师之所以只能评显露态，是因为发生过程是不可见的。这个限制现在也第一次松动了。同一件事既拆掉了旧的评价对象，也第一次让新的评价对象有了可见的可能——蒸发和显影是同一场事故的两面，取决于我们把装置造成什么样子。

所以真正的问题从来不是"怎么防住学生"。真正的问题是：

当那个不可外包的核第一次被逼到台面上，我们有没有胆量承认，我们过去两千年一直在评的，其实不是它。