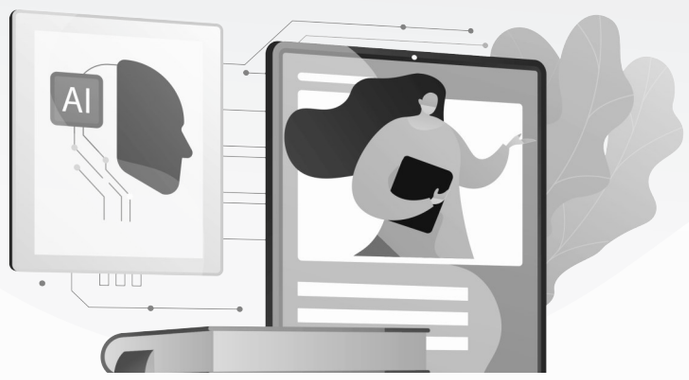
特别策划

人工智能何以赋能教学全过程

教育部组织的人工智能赋能教师发展现场会的召开，标志着教育数字化深化应用进入了一个新阶段。我们了解到，中央民族大学附属中学正在探索将人工智能全链条融合进课堂教学。该校副校长、中央民族大学教育学院副院长蔡可认为，人工智能是重要的学习工具，其优势应充分释放，而它固有的局限则有赖教师以专业判断力加以补足与校准。他们的具体做法有哪些？本报特别组织了本期专版，以飨读者。



备课：请准备好自己的“判断”

□ 许崇娜

不少教师都有过这样的经历：备课卡壳时，向人工智能发一句指令，几秒钟一份要素齐全的教案就生成了。教学目标、重难点、教学流程、板书与配套练习一应俱全，效率提升显而易见。但一所学校、一个班级、一个学段构成的教学情境带着独有的学情约束，凭什么能在几秒内被完整生成？

看起来合格的教学设计未必适配真实课堂。通用模型没接触过这些学生，不知道他们会在哪一处思维节点卡住，不知道上一节课留下了什么节奏缺口，更预判不出学生给出意外回答后，课堂该在哪里接续。它能把重难点排得整齐，却抓不住这节课真正的认知卡点，能把环节理顺，却拿捏不住课堂的松紧。

还有隐蔽的风险在于“幻觉”。一份文本排版规整、语句通顺，读者容易默认其逻辑已经过推敲，从而在不知不觉间跳过核验与思辨。教学设计真正费心的是持续的判断，而非环节的串联：这个环节为何走这条路径而不是另一条？学生答不出来，追问该往哪个方向？甚至这个环节本身要不要保留？工具能给出第一版方案，但往后关乎课堂走向的取舍只能由教师完成。

教学设计是多轮试出来的，不是一次写出来的。改稿和基于实践的反思才是设计的核心：一处环节时长被压缩，后面就跟不上；一道题太浅，学生在课堂上游神；一段材料的呈现顺序换了，理解的难易就完全不是一回事。只有经历过这几轮“不对劲”，才知道一节课为什么是这个样子。而人工智能给出的第一稿恰恰把这几轮都绕过去了。

面对人工智能，常见的应对是两头各走一头。青年教师容易对工具产生依赖：从教之初就能拿到成套资源，自己从零搭建教学设计的动力自然减弱。少了独立构思、推翻、重构这些环节，专业成长就少了一个载体。老教师的反应则相反，觉得人工智能生成的教案空话多、套路重，索性全面否定。

全盘依赖与全面否定，本质上是同一种回避，前者交出的是“判断”，后者交出的是“机会”。人工智能在备课链条里的位置应当是助手。素材检索、基础概念整理、格式规整这类事务性工作交给工具，省下的时间用来打磨贴合校本、班本内容。教学目标怎么定，重难点怎么判，问题链怎么搭，这些要由教师拿主意。分界线不难划：出了错只是格式难看，可以交回工具；出了错会打乱整节课的推进节奏，就得自己扛。

人工智能生成的内容常伴有事实错误或逻辑矛盾，要逐条核对。更关键的是，要依据本班学情评估其适配度：难度是否超出学生水平？预设是否符合学生习惯？经过这一轮核验，通用模板才能转化为面向本班的教学设计。带着学情标尺去校准AI初稿，对青年教师而言，反而是更实在的训练。

技术带来效率，也埋下专业能力钝化的风险。问题不在于是否使用工具，而在于是否将本该属于自己的思辨与判断一并外包。只有先厘清教学逻辑与学情边界再借助工具，效率才不会以牺牲判断力为代价。

命题：让人工智能真正读懂“难度”

□ 张晨颖

现在不少教师直接给指令，让AI围绕同一知识点分层生成基础、提高、挑战三类试题，试卷梯度很容易搭建。但分层命题是否贴合学情，指令怎么写是次要的，关键在于教师能否将学情判断清晰、准确传递给AI。

难度并非题目自带的属性，而是源于教师在上课、批改作业中对学生的了解：哪些题能做、大概能做对几道、卡在哪一步、哪种错误最常见。AI对难度的判断基于文本统计，比如哪些题被标记为难题、常出现在压轴位置。这些规律本身没错，但它指向的是“这类题”的平均表现。同一道题，有的班2/3的人能做对，另一个班可能只有一半人够得着；同一条难度指令，落到不同学校就是两回事。让一个没见过自己学生的工具去做一件依赖“自己的学生”的事，只能得到平均数，而平均数对谁都不完全合适。工具并非不努力，只是拿不到最关键的信息。

进阶路径也是如此。一个知识点在整个知识体系中的位置与前后概念的衔接及递进次序，体现了素养整合性的特点，教师对此心中有数；而工具虽能识别知识点，却未必清楚它们之间的层级与联系。直接让工具出“挑战题”，生成的题目很可能碎片化。

此外，教师自身也存在难点。成熟教师判断难度靠的是积累的经验，无须量化标准。但要把这份隐性经验转化为能说清的标准传给AI，就会发现很难开口。难度是以本班学生为标尺形成的感受，脱离这个参照，标准就落不了地。教师能凭经验拿捏一个班的情况，却很难将标尺提炼成稳定、可转述的规则。

一旦完成这套转化，命题质量会明显改善。教师先拿出自己打磨过、经课堂检验的“锚题”，标注每道题的达成比例、考查指向、推理步数及预判错点。有了参照样本，模型出的题才有依据，才不至于漫无目的地飘。肯拿出“锚题”的教师，等于把平时看不见的判断摆上台面；青年教师接手“锚题”，也能顺带学习老教师如何拿捏难度，还可以要求工具在出题时说明难度判断、关键步骤及最容易想岔的地方；说不清难在哪里的题，通常不该直接用，若工具的说法与自己的判断对不上，就按照自己的来。这动作花不了几分钟，却能把“人工智能出的题”变成“人工智能出、教师校准过的题”。

教师的专业判断无法像数据那样简单叠加，却能在一次次命题与校准中得以梳理、沉淀和传递。批改也是如此，讲评的方式、追问的角度都需要教师重新思考、重新表达，才能被他人借鉴。命题亦然，花在命题上的功夫往往远超出一道题本身。久而久之，还会带来额外收益：学生手中逐渐积累起一批经过验证的题目，大家清楚哪道题适合夯实基础，哪道题能激发思考。此时再借助工具，出题门槛降低，省下的时间便可双向利用：向上多做情境化改编，向下为掉队的学生单独选题。这些工作，工具是无法替代人的。

如今采集学情数据的条件也比过去完备得多，但AI生成的分层试题是否适配，最终仍要依靠教师对本班学情的判断进行校准，算法无法自动界定难度。分层命题并非将同一套“标准难度”的题目统一发给所有人，而是让每道题都贴近学生当下的实际状态。

学情分析：让诊断落到实处

□ 孙兆怡

学校大多不缺学情分析系统。哪个知识点全班掌握较差、哪类错误反复出现、哪些学生落后，报告比人工统计细致得多。但常见的场景是：教研会上一起看报告，点头认可，散会后备课本上写的还是老一套。已识别出的知识漏洞未转化成针对性练习，课堂节奏也未调整。报告做得再全面，若课堂教学照旧，便失去了意义。

问题出在“诊断”与“处方”之间。系统擅长描述，能显示全班在某知识点上的掌握比例，但教师需要的是“接下来这两节课该怎么上”的具体指导，“错了20%”这个数字无法直接提供答案。从数字到行动，中间依靠的是教师的专业判断，判断的效率决定了有多少学生能得到帮助。

时序是一道坎。课上完、作业批完后数据才能出来，此时教学窗口期往往已过。该知识点已翻篇，下一课开始讲新内容。分析虽准，但为时已晚。晚到的诊断只能纠正已经形成的漏洞，而无法干预正在产生的问题。

能力是另一道坎。系统能看出学生错在哪，但“处方”属于另一类决策：这道题明天该不该重讲？这个概念要不要换个例子？这组练习该怎么分层？这需要同时考虑教材、班级实际和多年教学经验，数据分析师做不了，系统也算不出来。

不少学校陷入了这样的循环：平台建了，培训开展了，报告也看得很勤，但教学改进却不多。原因在于报告本身不指向行动，这与重视程度无关。报告就像体检单，不是治疗方案，只告诉你“哪里有问题”，不告诉你“明天怎么办”。如果一份材料不能转化为三道题、两节课或一段辅导，就不会被真正用起来。

还有一种误区，是把“分层”理解成“分等”：给全班贴固定标签，反应快的做难题，慢的做基础题。这看似公平，其实与诊断关系不大。同一个学生，上学期可能需要基础练习，这学期可能需要挑战题。若按标签走，快慢两组学生都会被留在原地。只有跟着诊断走，分层才有意义。

教师拿出具体行动是有“成本”的。分层作业听起来只是把作业分成3份，实际成本比想象中大，最省事的选择永远是照旧。教师想改，仅有意愿不够，还得有能落地的办法。

办法其实不难找。省力的做法是把标准降一点：一共出四道题，一道所有人做，一道给掉队的学生，一道给中间水平的学生，一道给走得快的，谁做哪一道，依据上一次的诊断来定。第一轮不必精细，做上两三轮，哪些学生掉在哪类题上自然清楚。方案精不精细不是关键，先跑起来才要紧。

数据素养的核心，是把数据变成教学动作，看懂图表只是第一步。读懂平均分、认出掉队学生，再配上对应的学科实践（这是另一个要讨论的复杂问题），中间的分析、归因和匹配是最容易被略过的专业工作。诊断的价值，报告本身说了不算，要看之后有没有发生什么：学生知道自己错在哪，下一节课老师换了哪个例子，家长看到卷子配合做什么——这三件事都发生了，诊断才算完成。

技术让“发现”变得比过去任何时候都容易，但如何抉择、如何干预，仍是教师自己的事。读完报表，不要止步于第一页的汇总图表。

阅卷：教师应把时间用在学生身上

□ 高 冲

对善于使用人工智能的学校来说，将机械性的批改交给系统已是常态，这本身是进步。值得讨论的是它背后的问题：腾出来的时间，该往哪里去？

先说人工智能阅卷可能带来的弊端。批改本是教师了解学生的渠道，包括作业上的笔迹、涂改、犹豫和用力程度。人工智能接管后，教师就少了一个低成本认识学生的机会。批改在教师那里从来不只是判对错，还包括一次次具体、有针对性的回应：“这道题为什么错在这儿”“你上次问过的那道题和这道是一回事”“你那句话我读了3遍才看明白想说什么”，这些话无法从模板生成。系统能说“此处表达不清”，却说不出针对这个学生、让他愿意改的那一句。教师的回应里装着判断、鼓励和分寸，这是系统做不出来的。

省下来的时间最该拿去弥补这些缺失。有三个地方最该去。课前，看看哪方面内容全班都有问题，提前想好下一节课怎么补，把时间花在最容易出问题的环节，而不是平均用力讲完。课后，找那几个反复出错的学生聊一聊，系统能告诉他错在第三步，却聊不出他们为什么在那里停住。集体备课时，把这一周的错误类型理一理，看是教法的问题、讲解不够清楚，还是练习设计有问题。

这三个地方有个共同点，都是人对人做的事。学生不是一组分数，是几十个具体、有差异的人。批改效率提上来之后，师生的接触应该更多，而不是更少。先把学生的真实状况弄清楚，再把力气用在别人替不了的那一段上。

还有一个变化值得留意。批改是一件有节奏的工作，它把一周或一个单元框成几段：讲新课、收作业、批改、讲评。这套节奏被系统分走一部分后，教师很容易一节课接着一节课往下走，不停下来看看这几天的效果。数据其实已经把信号传递过来了——哪类错误连续出现，哪几个学生反复掉队。过去批改这一环占的份额太大，大到遮住了别的环节。现在这一环松开了，别的环节就该露出来：讲评讲得更准一点，辅导跟得更紧一点，作业布置得更对症一点。

“人工智能批改之后教师该做什么”问的不是技术，而是原来那段时间到底该用来干什么。批改只是把它填满了。现在它空了，正好可以想清楚。原本被批改占掉的那些时间，够办成不少事。

还想提一个多数教师没注意到的用法。省下来的时间如果只用在单个学生身上，收益其实有限：一个班总有那么几个掉队的，更普遍的问题出在中间那一段——全班都在某个地方磕了一下，只是程度不同。把这部分时间花在集体讲评的重新设计上，效果比逐个补更明显。一道题换个问法，一处材料换个顺序，一段讲解挪到下一节先讲，看上去只是动了动位置，实际上是针对集体的一次“处方”。

工具接管了低阶劳动，教师的素养就需要往高处走。这是一个具体问题：原本被批改占满的时间，现在该用来做什么？这个问题的答案比会不会用工具重要得多。

组卷：用好考试数据

□ 张雨石

一次考试提供的信息往往比多数人预想的要多。考后，平台会给出每道题的得分率、平均分、难度和区分度，教师据此进行讲评，这套流程已驾轻就熟。难点在于再进一步：如何根据考试暴露的问题，决定下一份试卷的关注点及考查方式。数据虽在，若试卷依旧老样子，这段“从数据到决策”的路径，正是用好人工智能的关键。

理想状态下，考后系统应自动生成分析，显示下一份试卷要在哪些板块聚力、学生欠缺什么，教师只需要负责判断与取舍。一次，我们学科组尝试智能组卷，问题恰恰出在反馈断层。命题工具按内容标签和难度要求生成试卷，结果成绩分布中间层扎堆。教师排查单题发现，两道大题的区分度几乎为零：高分与低分学生的表现相差无几。这类题目无法有效区分学生水平。

教研会上，教师找到了原因：“这两道题太绕了”“题干里的条件给得太多”。可惜，这些判断仅停留在口头，随即消散。下一份卷子生成时，系统仍按照原模板出题，学生依然做不动，数据依旧不理想。问题在于系统不知道上次发生了什么，这与计算能力无关。数据经过研讨却未记录，下次自然无法调取，改进链条就此断裂。

要接上这一环，首先要看清一件事：考试数据中最有价值的并非平均分和难度系数等“漂亮数字”，而是错误的具体形态。学生是概念不清、审题漏条件、计算方向错误，还是根本无从下手？不同的错误类型，对应不同的补救策略。上次那份卷子之所以无效，正是因为有一道题堆砌了过多要兼顾的条件，超出了学生的处理能力，而这在单纯的得分率上难以体现。

要接上这一环，关键在于教师先把判断记录下来。多数教研组都做过类似总结，但若仅止步于此，出卷时往往还是会回到凭感觉的老路。真正的改进必须落实到具体动作上，指望工具自动变聪明是不现实的，比如下一份试卷提高某类题的比例，或者去掉超纲题中的某个条件，将其换成更贴近生活的情境。

数据本身也有门槛。有的平台只提供一张表和一个折线图，有的虽给出了难度系数和区分度，但指标含义并不清晰。工具负责计算数据，而数据代表什么、是否合理仍需教师解读。此外，时间也是现实制约：哪一块该加力、哪道题该改都要先想清楚，而这很耗时。时间不够时，最省事的选择是按老套路出题。更根本的问题在于，日常考试常被沿用终结性考试的选拔逻辑，导致数据偏向筛选。建议先从最好用的部分入手：重点关注哪些题全班都错、错法是否一致。

让链条转起来，依靠的是具备数据解读意识的教师。这更像是一种习惯，而非强制机制：每次考完，挑一道错误最多的题认真分析错因，并记入备课本。记录的内容还要能衔接后续教学：这类错误下次要加强练习，那道题要换个问法。只有写的人心里有数，出卷的人才能用得上。

考试的价值在于向师生呈现真实的学习状态。卷面反馈比零散问卷更可信，但表格里的曲线不会自动改变教学。只有将数据转化为几道新试题、一次有针对性的重讲或一段相应的辅导，这次考试才算完成了诊断。下一次智能组卷能否更精准，关键在于是否有人愿意把观察到的细节记录下来。

学生：如何用好人工智能助手

□ 王亚兰 刘红昆

笔者所在学校做过一次小规模问卷：近3成学生承认，使用人工智能时最常用的方式是“直接获取完整解题步骤和答案”；超过一半的学生认同“长期直接看答案会让自己懒得思考”；还有学生说，脱离工具面对难题时，第一反应不是“我试试”，而是烦躁、想放弃。虽然样本有限，但这3组情况指向同一件事：问题不在个别学生，而在工具的用法上。

用得对不对取决于使用前后学生做了什么。调研还发现一处错位：绝大多数学生认可“先独立思考、再求助工具”，但真正做到的少之又少。原因是工具太方便——自己憋5分钟没结果，问一下10秒就有答案。这种依赖已经深入能力层面：从工具拿到的答案，在脑子里没有形成知识网络。它填补的是“知道”，而非“学会”。

核心在于工具的机制。工具提供即时反馈，旨在“快速拿到结果”，而学习恰恰不在这个环节。学习发生在人“卡住”的时候：找不到方法、想换思路、怀疑自己的判断。工具把“答案”接走，同时也接走了“卡住”这一环。在难题中反复试探、思索10分钟、试错两条路最后自己摸出一条可行路径，这段试错的过程才是理解知识的骨架。工具的便利恰恰剪断了试错的路径。评判工具价值的标尺从来不是用不用，而是用完之后学生留下了什么。

另一重损耗影响更大，即提问的动力被消解。习惯于直接索取答案后，学生很少再追问：错误的根源是什么？这个思路是怎么推出来的？条件变了，方法还成立吗？答案越来越容易拿到，自己生出的问题却越来越少。纸面成绩未必下滑，但思考的脉络会慢慢模糊。

针对这些情况，可以颠倒学习次序。旧流程是先做题、卡壳后看答案、看懂即结束；新流程则是借助工具理解思路后，脱离工具独立重做一遍，并标注最初卡住的节点，将“复制答案”转化为“自主复现”。

用对了工具的样子，教师是能看见的：学生的问题变多，也更具体。原来问“这题怎么做”，后来问“这一步为什么不能那样走”。答案来得快，他们就把省下的时间用在追问上，同样一节课能多问两三个“为什么”。探究能力长不长，要看学生的问题有没有变多。

作业设计是教师可以把握的关键抓手。针对同一知识点，作业既可以独立完成，也可以求助工具，但要附加一个条件：标出卡点，并在求助后独立复现。后者虽然耗时更长，但学生走过的思维路径完全不同。很多被归因为“自制力不足”的现象，其实是任务设计在引导学生走捷径。

因此，关键在于谁主导提问，禁不禁用工具反而是次要的。学生能否持续产生自己的问题，是区分工具使用好坏的分水岭。提问习惯的养成依赖日常积累：课堂上暂缓给出定论，先呈现学生已有思考；遇到错题，交由学生辨析成因；读完文本，先收集学生想追问的地方。这些细微的教学选择有的能涵养发问动力，有的则在消耗它。

人工智能让答案变得廉价，我们要警惕的是问题随之变得稀缺。应对这种变化最好的方式不是隔绝工具，而是引导学生在用完工具后，带着新的疑问继续探究。只要还能生出新问题，工具的使用就仍然属于学习过程。