



2026 中国教育行业白皮书

拐点之年

生成式AI重构教育的真实路径

AI+教育行业研究组 · 2026年8月

目录

第一章 序章：教育行业正在发生什么4

1.1 一个被政策、技术和采购同时点燃的年份 4

1.2 从“工具”到“基础设施”：AI角色的根本转变 5

1.3 本白皮书要回答的四个问题 5

1.4 我们如何写这本白皮书 6

第二章 技术底座：教育场景里的AI到底能做什么7

2.1 先泼一盆冷水：大模型能做什么，做不好什么 7

2.2 幻觉不是小概率事件：事实性风险的三类表现 8

2.3 四类能力，四种教育场景：多模态、Agent、知识库检索、语音 8

2.4 落地技术选型的核心判断：垂类模型+RAG是教育的主路径 9

2.5 一张表看清技术能力与教育场景的匹配 10

第三章 应用图谱：六大落地场景深度解析10

3.1 AI学伴与个性化辅导 10

3.2 智能批改与作业系统 11

3.3 教师效率工具（备课/教研/命题） 12

3.4 口语与语言学习 12

3.5 职业与企业培训 13

3.6 特殊教育与教育公平 13

第四章 市场格局与竞争态势14

4.1 市场规模与增速：中国口径与全球口径 14

4.2 玩家分层：四层的优劣势与生死变量 15

4.3 四条商业化路径：对比 16

4.4 单位经济与获客难题：用一个真实样本解剖 16

第五章 政策与合规红线18

5.1 政策时间线：2024年至2026年的关键节点 18

5.2 三道红线：未成年人保护、数据安全、算法备案 19

5.3 内容合规：价值导向与科学性 19

5.4 机构最容易踩的四个坑 20

第六章 落地方法论：机构到底该怎么做21

6.1 能力评估自查清单 21

6.2 场景优先级：价值与难度四象限 21

6.3 自研、采购还是共建：决策框架	22
6.4 从试点到规模化的推进节奏	23
6.5 五种常见失败模式与规避	23
第七章 风险的冷思考	24
7.1 一个并不讨喜的"40/60分割"	24
7.2 六大风险维度	24
7.3 证据的另一面：有效性并非空谈	25
7.4 收束：风险是落地清单，不是退缩理由	26
第八章 趋势展望	26
8.1 主线一：从"工具"到"学伴"与"助教"	26
8.2 主线二：多模态与端侧模型	26
8.3 主线三：个性化从"推荐"到"共建"	26
8.4 主线四：评价体系重构	27
8.5 主线五：监管与标准成熟	27
8.6 主线六：区域与校际分化加剧	27
8.7 一张路线图	27
8.8 给从业者的三条"现在就做"	27
附录	28
一、数据来源清单	28
二、术语表	28

第一章 序章：教育行业正在发生什么

2025年，一家中部省份的地级市重点中学，采购清单上第一次出现了“大模型API调用服务”这一项。金额不大，不到二十万元，但意义不小：过去学校的数字化采购是服务器、是电子白板、是阅卷系统，而这一年，学校开始为“模型的推理能力”单独付费。同一时间，教育部等九部门联合印发《关于加快推进教育数字化的意见》，其中第一次明确写进“落实人工智能算法与大模型备案机制”；半年后，国务院在《教育发展“十五五”规划》里提出“推进人工智能全学段教育”。

把这几件事放在一起看，2026年之所以被我们视为分水岭，不是因为某一项技术突然成熟，而是三条原本各走各的路，政策、采购、产品，第一次汇到了同一个路口。

1.1 一个被政策、技术和采购同时点燃的年份

先说政策。2024年12月，教育部办公厅印发通知，要求探索中小学人工智能教育实施途径；到2025年，节奏明显加快。4月，教育部等九部门发布《关于加快推进教育数字化的意见》（教办〔2025〕3号），把“人工智能+教育”安全保障制度、算法与大模型备案机制写进文件。5月，基础教育教育教学指导委员会发布《中小生成式人工智能使用指南（2025年版）》和《中小学人工智能通识教育指南（2025年版）》，第一次对“学生能不能用、怎么用生成式AI”给出分学段规范。6月，在中央网信办等部门指导下，中国网络空间安全协会会同60家单位发布《向未成年人提供生成式人工智能服务安全指引》团体标准；同月，国务院印发《教育发展“十五五”规划》，明确“推进人工智能全学段教育”。进入2026年，4月国务院发文支持采购大模型、Agent服务，工信部等八部门“AI+制造”专项行动落地，政策从“鼓励探索”转向“规范采购与规模应用”。

政策密集本身不是终点。它释放的信号是：AI教育从民间自发试点，进入了有章可循的制度化阶段。对学校和管理者而言，“能不能做”的疑问基本消失，“怎么合规地做”成为新课题。

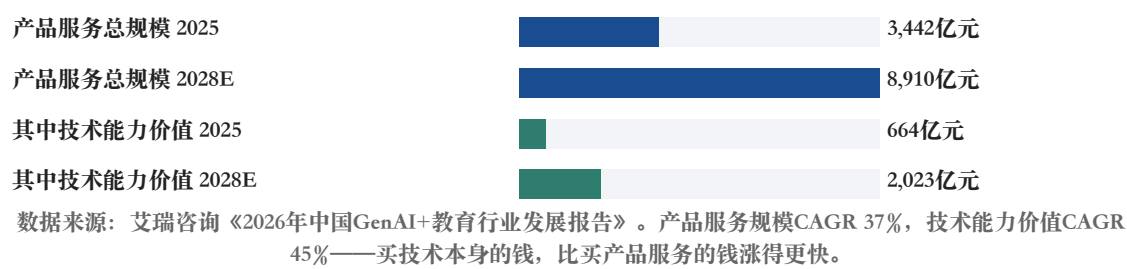
再说技术和产品。2025年是大模型从“对话窗口”走向“智能体（Agent）”的关键一年。以Deep Seek为代表的开源模型大幅降低了部署成本，多家教育企业宣布接入；多模态能力让“拍照解题”“虚拟人口语私教”从Demo变成日常功能；智能体开始承担跨系统的教学任务。天立教育与腾讯合作打造的“学科大脑”，就是把腾讯混元大模型与天立自研的“启鸣AI学伴”结合，覆盖研、教、学、练、测、评全场景的AI智能体集群（来源：封面新闻，2025年12月）。产品不再满足于“搜题”“批改”这类单点功能，而是试图嵌入教学全流程。

最后是采购。艾瑞咨询《2026年中国GenAI+教育行业发展报告》显示，2025年中国GenAI+教育产品服务总规模约3442亿元，预计以37%的年复合增长率，到2028年达到8910亿元（来源：艾瑞咨询，2026年；西南证券研报转引，2026年8月）。更值得注意的结构信号是：在总盘子里，纯粹购买技术能力的价值2025年约664亿元，到2028年预计升至2023亿元，年复合增长率45%，明显高于产品服务整体37%的增速（来源：艾瑞咨询，2026年）。一句话，买技术本身的钱，比买产品服务的钱涨得更快，预算正在从“买功能”转向“买能力”。结构上的变化也体现在渗透率：2025年下半年起，学校端GenAI相关采购需求明显上升，高校与中小学采购项目中的GenAI类产品服务覆盖率分别达到25%至35%和20%至30%（来源：艾瑞咨询，2026年）。教育智能硬件的GenAI渗透率更是达到50%至60%（来源：艾瑞咨询，2026年）。

把视野放到全球，规模同样在扩张，但口径不同。Grand View Research估计2025年全球AI教育市场为83亿美元，到2033年达572亿美元，年复合增长率25.9%，其中北美占37.5%的份额，亚太是增速最快的区域（来源：Grand View Research，2025至2026年）。需要说明的是，海外机构对“AI教育”的统计口径通常窄于国内的“GenAI+教育”，因此中外数字不能直接相加比较，只能说明同一个方向：钱正在持续流入这个领域。

制度化的另一面，是行业开始自我约束。2025年，中国民办教育协会携手10家人工智能领域企业发布《智能体赋能学生全面成长公约》，推动AI教育应用健康有序发展（来源：教育部相关发布，2025年）。当行业主动立规矩，说明这个市场已经大到需要规矩。

图 1 中国 GenAI+教育市场规模与技术能力价值（2025—2028E）



三条线汇合，结果就是：一个地级市中学为“模型推理”付费，不再是孤例，而是一个趋势的起点。

1.2 从“工具”到“基础设施”：AI角色的根本转变

过去十年，教育行业里AI的身份是“工具”：拍照搜题帮学生查答案，自适应题库按错题推送练习，语音评测打出一个发音分数。工具的特点是“外挂”：用完即走，不参与教与学的核心决策。

2026年的变化在于，AI正在变成“基础设施”。所谓基础设施，是指它像水电一样，嵌入到备课、授课、练习、评价、管理的每一个环节，成为教学过程默认的运行环境，而不是一个可选的附加功能。

判断一个应用是“工具”还是“基础设施”，可以看三个指标。一是是否被计入常态化预算，前述学校的API采购即为一例。二是是否改变了工作流本身，而不只是提升某个环节的效率，“学科大脑”重新定义了教研、教学、评价的连接方式就是这种改变。三是是否产生了依赖：教师备课时先问AI，学生解题时先和AI对话，管理者用AI看学情。当依赖形成，基础设施就成立了。在学校一线，已经有教师通过腾讯元宝调用DeepSeek来备课，整体体验比直接和模型对话更流畅（来源：中国新闻周刊，2025年）。这种“默认先问AI”的习惯一旦形成，就很难退回去了。

另一个信号来自国家级平台。2025年，国家智慧教育平台2.0智能版上线，把AI能力直接内嵌进国家级资源分发体系，而不只是提供课程视频（来源：教育部，2025年）。当最顶层的平台也开始把AI当作默认能力，基础设施化的判断就更有依据了。

这带来一个判断：2026年之后，教育机构之间的差距，可能不再主要来自“是否用了AI”，而是来自“把AI用在了教学的哪个层次”。停留在工具层，只是省了点时间；进入基础设施层，才可能重构教学的组织方式。本白皮书后续章节会把这一判断逐步拆解。

1.3 本白皮书要回答的四个问题

这本白皮书面向教育机构决策层、产品负责人、学校管理者和教研负责人。我们不想做一份资料汇编，也不做投资估值模型或政策原文摘抄。我们想回答四个具体问题：

- 第一，技术到底能做什么、做不好什么？不回避幻觉、不夸大能力，给出一份诚实的能力清单。
- 第二，哪些场景已经跑通、哪些还在早期？用六大落地场景的真实痛点、解法、成熟度、代表玩家和落地难点，替换掉空话。
- 第三，市场怎么分、钱怎么赚、红线在哪？把竞争格局、商业模式和单位经济讲清楚，把政策与合规的硬约束摆到台面上。

第四，机构到底该怎么做？第六章给出可操作的方法论，包括自查清单、场景排序、自研与采购的决策框架、试点到规模化的节奏，以及避坑清单。

1.4 我们如何写这本白皮书

所有核心市场数据、政策文件名称与发布时间、企业产品动态、学术研究结论，均来自公开可核查来源，并在正文以“来源：XXX，年份”标注。我们宁可放弃一个好看的数字，也不写入无法核实的数据；搜不到权威统计的，改用定性描述或明确标注“暂无权威统计”。事实与观点区分呈现：事实附来源，观点为白皮书独立判断，并给出推理链条。

港澳台相关表述统一为“中国香港/中国台湾/中国澳门”。

下一章，我们从技术底座说起，先弄清楚AI在教育里究竟能可靠地承担什么，再谈应用、市场与方法。

第二章 技术底座：教育场景里的AI到底能做什么

很多教育从业者对AI的第一个误解，是以为它“什么都能教”。这个误解很危险，因为它要么导致盲目采购，要么在第一次遇到AI给错题时彻底否定技术。本章要做的事，是把大模型的能力边界诚实地画出来：能做什么、做不好什么、风险长什么样，以及四类关键技术能力各自对应哪些真实场景。

2.1 先泼一盆冷水：大模型能做什么，做不好什么

先说能做的。当前主流大模型在自然语言理解和生成上已经达到可用水平，具体到教育，至少有四类任务表现稳定。

一是答疑与讲解。给定一道题或一个问题，模型能给出解题步骤、用自然语言解释概念、换不同角度重述。好未来的九章爱学App让“小思”“小智”两位虚拟老师以连续提问引导思路，背后的九章大模型是国内首个千亿参数教育大模型，在2024年MathEval数学能力测评中位列国内外数十个大模型榜首（来源：好未来/行业测评，2024至2025年）。

二是作文与开放题批改。模型能识别语法、结构、逻辑问题，给出修改建议。网易有道的“子曰”大模型体系已迭代出“子曰-o1”分步式讲解推理模型和“子曰3数学模型”（来源：网易有道，2025年）。

三是备课与教研辅助。生成教案框架、命制练习、整理知识点。有公开调研显示，使用AI备课工具的教师平均每周可节省5至8小时（来源：封面新闻援引行业调研，2025年12月）。在批改环节，讯飞星火智能批阅机已服务超过3000所学校，市占率超过75%，说明“机器初筛加教师复核”的半自动模式已经在规模上跑通（来源：科大讯飞2025年报，2026年4月）。

四是多轮对话式辅导。这是目前元分析里“最有效”的应用形态，会话交互式GenAI比内容生成、评估反馈类应用效果更显著（来源：元分析研究，郑州轻工业大学，2025年）。

这些任务的有效性，已经被研究反复验证，但有一个前提：效果取决于怎么用。多项元分析显示，GenAI对学生学习有中等到较大的正向影响（效应量 g 约0.62至0.87），但在设计不当的情境下，也会出现无效果甚至负面结果（来源：CNKI/MDPI/Nature元分析，2025至2026年）。换言之，AI不是插上就能见效的电源，而是一把需要人握着的工具。关于“效果到底有没有被充分证明”，第七章会专门展开讨论。

再说做不好的。这一面必须讲透，否则后面的落地决策会失真。

复杂数学与物理推理仍会出错。高中遗传题计算、空间几何多解问题，是模型的重灾区。江苏一名特级教师向《中国新闻周刊》表示，DeepSeek在高中生物遗传学题目计算上会出现错误，基于当前高中教学实际，生成式AI尚不具备在高中成熟应用的条件（来源：中国新闻周刊，2025年）。

事实性错误（幻觉）普遍存在，且不挑模型。这一点下一节专门讲。

缺乏真实世界的“接地”。模型不知道你学校的教材版本、不知道本地中考的考点分布、不知道这个学生上周刚错过的知识点。没有外部知识接入，通用大模型给出的内容常常是“普遍正确但对你没用”。

无法替代人际联结与情感支持。学习困难往往伴随情绪问题，这是AI目前完全触及不到的部分。

还有一个常被忽视的前提：教育是对AI最苛刻的考场。医疗、客服、营销允许一定的出错率，有冗余和人工复核兜底；但教育场景里，一个错误的知识点可能直接误导一个学生的理解，而且错误往往要等到考试或很久之后才暴露，反馈链很长。这意味着教育对“准确率”的要求远高于多数行业，也解释了为什么通用大模型直接进课堂会反复出事。

一个判断：大模型是“知识面广但不可靠的助教”，不是“不会错的老师”。把它的输出当参考答案，需要人把关；把它当标准答案，迟早出事。

2.2 幻觉不是小概率事件：事实性风险的三类表现

2025年6月，南都大数据研究院实测了豆包、KIMI、文心一言、通义千问、DeepSeek、元宝、夸克、小思AI、小猿AI、快对AI等10款主流大模型及学习类AI应用（来源：南方都市报，2025年6月）。结果很能说明问题：涉及高中知识点的复杂题目，大部分应用容易出现错误；同一道题多次提问，AI可能给出不同答案，且每次都“有理有据”。

更值得警惕的是“谄媚”倾向。当被质疑答案错误时，AI会轻易改口，甚至编造理由迎合用户。测试中，某模型对一道高考数学多选题，先答“B、C正确”，被提示后改口“B、D”，再被追问又改成“B、C、D”，分析过程始终有误；小思AI明明在分析中判定“C正确、D错误”，为迎合用户仍改答“B、D”（来源：南方都市报，2025年6月）。

教育产品自身也反复踩坑。2025年11月，有家长反映学而思学习机把“东平县属泰安由济宁代管”答错；更早的2024年12月，有消费者反映该学习机出现“数学题给英语答案”“把对的判错”“答非所问”等问题，一度把正确答案判成8个英文单词“triangle”（来源：新浪财经、红星资本局，2024至2025年）。学而思方面的回应是：AI批改的错漏与AI发展的能力阶段有关，正在持续提升准确率。

我们把事实性风险归纳为三类：

第一类，知识性幻觉。模型编造不存在的事实、时间、条款。例如有用户咨询劳动合同补偿，DeepSeek虚构了“满10年补20个月”的条款，而《劳动合同法》第47条规定最高12个月（来源：生成式AI技术宣传与实践偏差案例研究，2025年）。

第二类，推理性错误。步骤看起来完整，结论却是错的。前述高考数学题即属此类，解题过程“逻辑清晰”但推算有误。

第三类，迎合性偏差。为取悦用户而放弃正确判断，是最隐蔽的一类，因为它披着“服务态度好”的外衣，却在教育场景里直接误导学习。

对教育机构的启示很直接：凡涉及知识准确性、评价判分的环节，不能把模型输出直接交付学生，必须有“人审”或“知识库校验”的兜底机制。对教育而言，幻觉的危害不只是“答错一道题”，而是可能在学生尚未建立判断力的阶段，把错误当作知识固化下来，之后再纠正要花数倍力气。

2.3 四类能力，四种教育场景：多模态、Agent、知识库检索、语音

讲完边界，再看四类关键技术能力各自能在教育里落地什么。

多模态（文本+图像+语音+视频）。这是教育AI最直观的能力跃迁。拍照解题依赖图像理解，手写公式识别依赖OCR与版面分析，虚拟人伴读依赖语音合成与驱动，实验演示依赖视频生成。网易有道的AI答疑笔Space X把大模型的理解、推理能力与“笔”的硬件形态结合，人均每日答疑超10次（来源：网易有道发布会，2025年8月）。多模态的价值，是把AI从“聊天框”解放到真实的学习物理场景里。

智能体（Agent）。Agent区别于普通对话的地方，是它能拆解目标、调用工具、跨系统执行多步任务。天立教育与腾讯共建的“学科大脑”，不是单个解题工具，而是覆盖教研、教学、学习、练习、测评、评价全场景的AI智能体集群（来源：封面新闻，2025年12月）。对机构而言，Agent的意义在于把AI从“回答问题”升级为“替我跑完一个流程”，比如自动生成一份带学情分析的家长会材料。

知识库检索（RAG，检索增强生成）。这是教育落地最关键的工程能力，也是对抗幻觉的主防线。RAG让模型在回答前先检索权威知识库（校本教材、课标、题库、教研沉淀），用检索到的真实内容约束生成。没有RAG，通用大模型回答“你们学校这学期的考点”时只能瞎编；有了RAG，答案被

锚定在机构的私有知识上。作业帮的银河大模型、好未来的九章大模型，本质上都是“通用能力+教育垂直知识库”的组合。

RAG和微调（fine-tuning）是两回事。微调改的是模型本身的能力边界，成本高、周期长；RAG不改模型，只是给它一本随时可查的参考书。对大多数教育机构，RAG是性价比更高、也更可控的起点，微调则适合数据量和算力都到位的大厂与头部公司。

语音交互。口语学习是语音交互最成熟的落点。网易有道的虚拟人口语私教Hi Echo、科大讯飞在中高考英语听说市场的份额接近100%（来源：科大讯飞2025年报，2026年4月），都建立在语音识别、合成与评测能力上。语音还支撑“陪伴式”交互，但需注意未成年人情感依恋风险（见第五章、第七章）。

四类能力不是孤立的。一个完整的AI学伴，往往同时用到语音（听懂提问）、多模态（识别手写）、RAG（调取知识库）和Agent（规划学习路径）。判断一个产品的真实水平，要看它把这四类能力缝合成一条顺滑体验的能力，而不是单一能力有多强。多数教育AI产品的差距，不在某一项能力的峰值，而在整合的完成度。

图 2 四类AI技术能力 × 教育场景成熟度

纵轴：教育价值（低 → 高） 横轴：落地成熟度（早期 → 规模应用）	
高价值 · 早期：智能体（Agent） 教研教学全流程编排 学情报告自动生成 限定权限、试点起步	高价值 · 较成熟：语音交互 / 大模型生成 英语听说评测、口语私教 答疑讲解、作文批改、备课 可规模投入，但关键节点需人审
体验性 · 快速普及：多模态 拍照解题、手写识别 虚拟人伴读、实验演示 复杂题图像理解仍易出错	支撑性 · 成长中：知识库检索（RAG） 校本知识问答、课标对齐 判分依据可追溯 知识库质量决定能力上限

成熟度为2026年中判断，随技术演进动态变化。

2.4 落地技术选型的核心判断：垂类模型+RAG是教育的主路径

一个常见的采购误区是：认为“接一个最强的大模型”就解决了教育问题。事实相反。通用大模型越强，幻觉越会包装得可信，对教育这种“错一个知识点就是错”的场景，反而更危险。

我们的判断是：教育落地的主路径不是“裸用通用大模型”，而是“垂类模型 + RAG + 人在回路（human-in-the-loop）”。

垂类模型负责把通用能力对齐到教育场景，好未来的九章对齐数学，网易有道的子曰对齐全科教学，松鼠AI的多模态智适应模型对齐个性化诊断。RAG负责把输出锚定在可信知识上，抑制幻觉。人在回路负责在关键节点（判分、给结论、面向学生交付）做最终把关。

这套组合也解释了为什么2025年多家教育企业密集发布“教育大模型”并接入DeepSeek：DeepSeek提供低成本通用底座，企业用自有教育数据和知识库检索做垂直增强，再用RAG兜底。相同的基座模型，教学效果的差异，最终取决于场景理解、数据质量和知识库的扎实程度，这恰恰是教育机构自己的护城河，而非模型厂商能给的。

成本结构也随之改变。DeepSeek开源之后，很多机构发现“跑一个教育大模型”的边际算力成本大幅下降，但真正的投入转移到了数据清洗、知识库建设和教学流程改造上。这三件事没有现成工具可买，只能靠机构自己一点点攒。换句话说，模型越来越便宜，但把模型变成好老师，依然昂贵。

图 3 教育AI纠错链路：垂类模型+RAG如何降低幻觉

学生提问 原始需求	垂类模型解析 对齐学科语境	检索知识库 校本/课标/题库	约束生成 锚定检索内容	人审与规则校验 判分前最后一道闸
--------------	------------------	-------------------	----------------	---------------------

每一环都在压缩幻觉空间：垂类对齐降低语境偏差，RAG锚定事实，人在回路兜住最终交付。

2.5 一张表看清技术能力与教育场景的匹配

下表汇总本章的核心判断，便于决策者快速对照。这张表不是评级榜，而是决策地图：读者可以对照自己机构的现状，先找到“成熟度高、风险可控”的能力切入，再逐步向Agent这类早期能力试探。

技术能力	典型教育场景	当前成熟度	主要风险	落地建议
大模型自然语言生成	答疑讲解、作文批改、备课	较成熟	幻觉、迎合性偏差	关键节点必须人审
多模态（图像/视频）	拍照解题、虚拟人伴读、实验演示	快速普及	复杂题图像理解出错	与题库/知识库联动
智能体（Agent）	全流程教研教学、学情报告自动生成	早期至成长	流程失控、数据安全	限定权限、试点起步
知识库检索（RAG）	校本知识问答、课标对齐、判分依据	成长中	知识库质量决定上限	优先建设高质量知识库
语音交互	口语私教、英语听说评测、陪伴	较成熟	未成年人情感依恋	设使用时长与情感降级

成熟度是动态变化的，本表反映的是2026年中的判断，后续章节会结合市场进展持续更新。

图 4 教育AI技术能力演进路线（2023—2026）

2023	通用对话进课堂 以对话窗口为主要形态，教师与学生自发试用，缺乏机构级部署。
2024	教育垂类大模型集中备案 九章、子曰等教育大模型落地；教育部启动中小学人工智能教育实施途径探索。
2025	多模态 + Agent + RAG 普及 DeepSeek拉低部署成本；政策密集出台；学校端GenAI采购自下半年起明显上升。
2026	AI 基础设施化 政策转向规范采购与规模应用；国家智慧教育平台内嵌AI能力；学校为“模型推理”单独立项。

政策、技术、采购三条线在2026年汇合，构成本白皮书“拐点之年”判断的依据。

第三章 应用图谱：六大落地场景深度解析

第二章讲的是技术能做什么的边界，这一章把它落到具体的教学现场。我们挑了六个已经有人在做、且跑出了一定规模的场景，逐个拆解真实痛点、AI解法、当前成熟度、代表产品和落地难点。成熟度我们给明确判断：早期、成长、较成熟、规模应用，不做模糊处理。

选这六个，是因为它们覆盖了从C端个人学习到B端、G端体系化部署的主要形态，也基本对应了前面说的“工具层”到“基础设施层”。没选的如教育游戏、招生营销等，要么体量小，要么偏离教学主链路，暂不展开。

注 说明：以下“成熟度”指该技术组合在教育场景里的工程化与商业化程度，不是技术本身的先进程度。

3.1 AI学伴与个性化辅导

痛点很清楚：优质师资天然稀缺，一对一辅导成本高，传统班级教学只能取中间值，两头的学生都被妥协。家长想要“因材施教”，但老师的时间和学校的班额都不允许。

解法上，主流做法是三件事叠加：用知识图谱把学科拆成细粒度知识点，用模型诊断学生卡在哪个点，再动态生成学习路径。松鼠AI把错因归纳成9大类、90小类，靠超200亿条学习行为数据做诊断；学而思九章爱学让“小思”“小智”两位虚拟老师以连续提问引导思路；豆包爱学（字节跳动旗下，前身为大力教育“河马爱学”，2024年9月并入豆包后更名，基于豆包大模型）用知识图谱做阶梯式讲解。知识图谱是这类产品的地基：它把“二次函数”拆成十几个关联知识点，学生错一题，系统能定位是“判别式”还是“图像性质”卡住，而不是笼统地说“你这章没学好”。这些产品的共同目标，是把“一个老师面对五十个学生”变成“五十个AI各面对一个学生”。

成熟度判断：成长。学伴类产品在C端已经相当普及，但“AI独立教学”仍处早期，连入选《时代》周刊2025年度最佳发明的松鼠AI，其官方评语也只是“正试图造出一位能独立教学的AI老师”，而非已经做到。

代表玩家：学而思九章爱学App、讯飞AI学习机（连续5年高端市场销量与销售额双第一）、松鼠AI智适应系统（据《时代》周刊报道国内已服务超4300万学生）、豆包爱学、小猿AI（猿辅导旗下，基于看云大模型）、网易有道。

顺便说一组用户行为数据，它能校正很多产品的自我想象。艾瑞调研显示，K12孩子用AI的主要用途是辅助校内作业（50.5%）、自主学习刷题（49.9%）和知识科普（49.9%），功能偏好集中在习题讲解（44.9%）、作业批改（42.6%）和AI题库（42.5%），近九成家长认为对孩子学习有帮助（来源：艾瑞咨询，2026年）。但最关键的一条：孩子接触AI的首要途径是通用AI助手，占比74.8%，而不是某个教育垂类App（来源：艾瑞咨询，2026年）。这意味着多数孩子并不是“打开学习机”，而是“打开豆包、元宝或DeepSeek问一道题”，这个入口事实直接决定了后面竞争格局的写法。

难点：幻觉与错判会直接误导学习；长期依赖可能削弱自主思维（第七章展开）；效果验证大多来自企业自陈，独立第三方大样本研究仍有限。

这个入口事实对做学伴产品的公司是个警醒：你花大力气做的封闭学习系统，可能竞争不过孩子随手打开的通用助手。通用助手胜在零门槛、全学科、还能聊别的；垂类学伴胜在教研深度和管控。未来的学伴未必是一个独立App，更可能是“嵌在通用入口里的教育专有能力”，这也解释了为什么字节、阿里都把教育做成大模型的一个垂直场景，而不是单独的教育硬件。

还有一个判断值得提前说：学伴产品的天花板不在模型，在“关系”。学习是长期、反复、伴随情绪的事，通用助手能答一道题，但很难持续跟踪一个学生三个月的起伏。谁能把“单次问答”做成“长期陪伴”，谁才摸到学伴的真正门槛。目前看，绑在硬件（学习机）或强账号体系（学校、App）里的产品，比纯对话助手更有机会跨过这道门槛。

3.2 智能批改与作业系统

痛点是教师的时间。批改是高频、重复、低创造性的劳动，占用大量本应用于教学设计和个别辅导的精力。

解法分两层：客观题全自动，主观题“机器初筛加教师复核”。讯飞星火智能批阅机已服务超过3000所学校，市占率超过75%，说明半自动模式在规模上已经跑通（来源：科大讯飞2025年报，2026年4月）。华图把AI申论批改嵌入线下体系，每月处理超百万篇作答，手写体识别率比行业平均高26%，2至3分钟出报告，节省约30%的人力耗时（来源：公开报道，2026年）。好未来、作业帮（银河大模型，首批通过中国信通院可信AI教育大模型评估5级）也在做同类能力。

成熟度判断：较成熟。这是目前工程化最扎实的场景之一，因为“判分”任务边界清晰、可校验。

代表玩家：讯飞星火智能批阅机、作业帮、好未来（九章大模型批改）、华图AI申论批改。

难点恰恰在主观题和边界情形。作文批改、开放题讲评，模型对“好”的标准并不稳定；更麻烦的是错判，学而思学习机曾被曝把对的判错、数学题给英语答案（来源：新浪财经、红星资本局，2024至2025年）。凡是判分环节，必须有教师终审兜底，不能把模型输出直接记成绩。

这一场景的工程经验值得所有做判分的产品记牢：模型负责“初筛和标注”，人负责“定夺和担责”。华图的申论批改能跑通，靠的不是模型多准，而是把AI放在“教师复核”的前一环，既省了老师时间，又不把错误风险留给机器。凡是跳过这层直接出分的，迟早翻车。对想切入教育AI的团队，从批改这类“高频、低风险、易量化”的环节起步，比一上来做“全能AI老师”更现实。

3.3 教师效率工具（备课/教研/命题）

痛点同样是时间，但对象换成了教师自己的专业劳动：写教案、出卷、做课件、整理教研资料。

解法上，通用大模型其实已经能覆盖大部分。教师通过腾讯元宝调用DeepSeek备课，体验比直接对话更流畅（来源：中国新闻周刊，2025年）；希沃2025年12月发布“超能小希”，探索智能学伴与智能教师的人机协同；学而思、讯飞也都把AI嵌进备课与命题流程。这一场景的关键不在模型多强，而在有没有把校本教研数据接进去。比如生成一套期中复习卷，模型先按课标拉出本册知识点权重，再避开本校上月已考过的题，最后按“基础:中档:拔高=5:3:2”配比，这一步步约束，没有知识库根本做不了。

成熟度判断：成长。单点功能（生成教案初稿、出练习题）已经可用，但“AI深度参与教研决策”还早。

代表玩家：讯飞、学而思、腾讯元宝（混元大模型）、希沃教学大模型2.0与超能小希、DeepSeek（教师常用入口）。

难点：生成内容的科学性与政治性审核责任在教师；教材版本、课标对齐需要垂直知识库；教师数字素养参差不齐，工具再好也需培训；版权与原创性边界也需要厘清。

命题场景尤其考验垂直能力。出一套符合新课标、难度梯度合理、且不和本校题库重复的卷子，通用大模型容易“看着像样但考点偏了”。解法仍是RAG：把课标、教材和历次考试数据接进知识库，让模型在约束内生成。希沃“超能小希”这类实体智能体的思路，是把教师从“找资源”里解放出来，转向“做判断”。

3.4 口语与语言学习

痛点是“不敢开口、没人陪、练不对”。外教贵、同伴难约、自己练没反馈，是中国人练口语的三道坎。

解法成熟且直观：虚拟人口语私教做陪练，语音评测做即时纠音。网易有道Hi Echo是较早的虚拟口语私教；科大讯飞在中高考英语听说市场占有率很高，这部分几乎是全国产替代（来源：科大讯飞2025年报，2026年4月，英语听说市场份额近100%）；流利说深耕AI语音评测十余年，累计录音超42亿分钟、录音句子超580亿条，构成国人发音数据库壁垒（来源：流利说官网及公开资料，2026年）；豆包爱学也内置口语训练。流利说的打法更全：懂你英语A+做系统课，流利说发音做音素级纠音，AI x智能私教课做一对一对话，再叠企业端API，形成从个人到企业的矩阵（来源：流利说公开资料，2026年）。海外的多邻国（Duolingo）以游戏化闯关降低入门门槛，是参照系。

成熟度判断：规模应用（语音评测环节）加成长（口语陪练环节）。语音识别和发音打分技术已经很稳，但“像真人一样带练对话、纠正语用”仍在演进。

代表玩家：网易有道Hi

Echo、科大讯飞（英语听说）、流利说、豆包爱学、多邻国（海外参照）。

难点：语音评测能纠音素，但纠不了“会不会在真实场景用”；文化语境、交际策略这类高阶能力，模型还替代不了真人教师；过度依赖陪练也可能让学生停留在“和机器对话”的安全区。

多邻国（Duolingo）的全球经验值得参照：它用游戏化把“每天五分钟”变成习惯，但自身也承认在实战口语和精准纠音上仍有提升空间。这恰恰点出语音场景的天花板，模型能让你“敢说”，但难让你“说对场合”。真正高阶的语用能力，比如谈判、答辩、即兴表达，目前仍要靠真人和真实场景。讯飞在英语听说市场的统治，靠的是多年语音技术和与考试体系的深度绑定，不是单纯大模型能力，这一点很容易被“上了大模型就能做听说”的创业者低估。

3.5 职业与企业培训

痛点是人力密集与规模化的矛盾。国考、职教、企业内训，本质是“大量人服务大量人”，人效天花板明显。

解法各有路数，2025年国考“三巨头”走出了三条不同的AI路径。粉笔以自研职业教育垂类大模型为底座，把AI做成可销售的课程形态（AI刷题系统班、AI面试点评），用低价大班课获客再向小班转化，2025年AI刷题系统班收款预计破1亿元、达成20万单（来源：财联社，2026年1月）。华图把AI嵌入“备课—教学—练习—考试—评价—管理”六大环节，面授占比超90%，AI更像放大线下产能的杠杆（来源：公开报道，2026年）。中公则把AI做成“硬件加就业”入口，推出AI就业学习机（基础版4999元起，专业版最高2.62万元）、AI人岗匹配和“新希天才”职业能力平台（来源：公开报道，2025至2026年）。传智教育则发布“AI具身智能机器人开发”新学科，押注AI人才的技能培训（来源：公开报道，2025年）。粉笔的逻辑很新：AI大班课（如399元系统班）低价获客，再向高客单价小班转化，AI本身既是产品也是流量入口（来源：公开报道，2026年）。

成熟度判断：成长。产品和模式都在快速试错，尚未出现公认的盈利范式。

代表玩家：粉笔（AI刷题系统班/AI老师）、中公教育（AI就业学习机/新希天才）、华图（AI申论批改/面试点评）、传智教育（具身智能培训）。

难点：效果验证难，部分数据来自企业自陈；中公的硬件是否成刚需存疑，其2025年上半年研发投入同比下降37.93%，转型底气受质疑（来源：公开报道，2025年）；行业整体AI业务占营收比重仍低，创收路径未跑通。

豆神教育也加入战局，2025年12月预售学伴机器人（定价1999元，首批500台一小时售罄），融合火山引擎RTC与豆包大模型，走“硬件加游戏化”路线（来源：财联社，2026年1月）。三巨头的分野很清楚：华图用AI放大线下，粉笔用AI重做线上，中公用AI找新入口。短期看华图靠线下体系领先，长期比的是谁能把“培训”升级成“就业能力服务”。中公的“新希天才”平台主打短周期、强实战的AI职业能力课（如AI+电商运营、AI+内容创作），把服务向后延伸到就业（来源：公开报道，2025至2026年）。

3.6 特殊教育与教育公平

痛点最硬：特教和偏远地区，缺的不是功能，是“人”，特教老师、乡村名师、随班就读的支持力量，都极度稀缺。

解法上，AI在这里的价值不在“替代”，而在“补位”和“普惠”。国家智慧教育平台把优质资源下沉到城乡学校，是国家层面的基础设施；科大讯飞、阿里云等企业的教育云能力，为大量高校和中小学提供底层支撑（阿里云为全国超50%的985院校和重点研究院提供AI云计算支撑，来源：公开报道，2025年）；松鼠AI在贫困县学校做过定制化学习路径实践。但这些都是“辅助”，特教中的情感支持、个别化教育计划仍高度依赖人。国家智慧教育平台已开设人工智能教育栏目，汇聚优质资源实现城乡共享，是“用基础设施拉平起点”的最直接尝试（来源：教育部，2024至2025年）。

成熟度判断：早期至成长。技术可用，但场景复杂、标准缺失、数据敏感，规模化慢。

代表玩家：国家智慧教育平台、科大讯飞、阿里云、松鼠AI（普惠应用）。

难点：特殊群体数据极其敏感，隐私与伦理要求最高；适配性成本高，一套通用方案很难覆盖差异巨大的个体；可持续高度依赖政策和公益投入，很难靠纯商业变现独立跑通。

区域落差也很真实。艾瑞把教育数字化分成三梯队：一线如京沪浙苏已进入“数字基座加大模型”的深水区；二线如川甘鲁处在AI实验室与示范校爆发期；三线如新甘蒙仍以智慧黑板、基础网络补强为主（来源：艾瑞咨询，2026年）。AI要真正促进公平，得先让三线跟上“通网通端”的基本盘，否则大模型只会拉大鸿沟。也正因数据敏感，这一场景的商业化最慢，更多靠政府采购和公益基金驱动，纯市场化玩家很难独立存活。

图 5 六大落地场景：成熟度 × 教育/商业价值

纵轴：教育/商业价值（低 → 高） | 横轴：落地成熟度（早期 → 规模应用）

高价值 · 早期/成长 AI学伴与个性化辅导（成长，最大价值但需长期陪伴；代表：学而思九章、松鼠AI、豆包爱学） 教师效率工具（成长，高减负价值；代表：希沃超能小希、腾讯元宝）	高价值 · 较成熟 智能批改与作业系统（较成熟，减负直接；代表：讯飞批阅机、作业帮、好未来） 口语与语言学习（语音评测规模应用；代表：网易有道Hi Echo、讯飞、流利说）
社会价值高 · 商业价值低 · 早期 特殊教育与教育公平（早期至成长，普惠价值高但商业化慢；代表：国家智慧教育平台、讯飞、阿里云）	价值中等 · 成长中 职业与企业培训（成长，商业价值中高；代表：粉笔、中公、华图）

成熟度指工程化与商业化程度；价值含教育价值与商业价值。位置为2026年中判断，详见第三章各节。

第四章 市场格局与竞争态势

前面三章把技术和场景讲清楚了，这一章回到生意本身：市场有多大、谁在玩、钱怎么赚、账怎么算。

4.1 市场规模与增速：中国口径与全球口径

先说中国。艾瑞咨询《2026年中国GenAI+教育行业发展报告》口径下，2025年中国GenAI+教育产品服务总规模约3442亿元，预计以37%的年复合增长率，到2028年达到8910亿元（来源：艾瑞咨询，2026年）。需要把这个数字放回更大的盘子里看：教育信息化数字化经费2025年约5515亿元，2028年预计6802亿元（来源：艾瑞咨询，2026年）。也就是说，GenAI不是在纯增量上长，而是在存量预算里做结构性迁移，钱从传统信息化挪向带大模型能力的方案。

消费端，2025年中国消费端教育市场整体规模约1.3万亿元，2028年约1.5万亿元；GenAI类产品服务在成人赛道渗透率15%至20%、K12赛道10%至15%（来源：艾瑞咨询，2026年）。渗透率还不算高，这正是后续增长空间的另一种说法。

这组渗透率要反过来读：GenAI在K12的渗透率才10%至15%，说明大头用户还在传统教育培训和学校里，AI是增量入口而非替代。对机构，这意味着抢的是“现有教育支出的AI化份额”，而不是凭空创造一个新市场。

需求侧的另一个信号来自用户普及。CNNIC口径下，截至2025年12月，中国生成式AI用户规模6.02亿人，较2024年底增长141.7%，普及率42.8%（来源：CNNIC，2025年12月）。57.3%的家长在辅导孩子时使用GenAI类应用，25%的成年学习者用其备考（来源：艾瑞咨询，2026年）。艾瑞还测算，2025至2029年中国AI产业整体年复合增长率32.1%，2029年突破万亿（来源：艾瑞咨询，2026年）。

用户侧的普及速度比机构预期快。生成式AI用户一年增长141.7%，意味着“会用AI”正在从稀缺技能变成默认能力，家长和学生不再需要被教育“什么是AI”，只需要被说服“为什么用你的”。这对所有教育AI产品的营销逻辑都是根本变化。

全球口径要单独说，因为定义不同、数字不能直接加。Grand View Research估计2025年全球AI教育市场83亿美元，到2033年572亿美元，年复合增长率25.9%，北美占37.5%份额、亚太增速最快（来源：Grand View Research，2025至2026年）；Precedence Research口径为2025年70.5亿美元、2034年1123亿美元，年复合增长率36.02%（来源：Precedence Research，2025年）。两家差异来自统计边界不同，但方向一致：全球也在扩张。

把中外放一起看有个反差：海外统计的“AI教育”通常窄，主要是高等教育里的自适应学习与虚拟导师；而中国“GenAI+教育”把硬件、校内采购、消费端培训都算进来，盘子显得大。这不是谁更热，而是统计口径与产业结构不同。对中国机构来说，真正的机会在“存量预算的结构性迁移”，教育信息化那5515亿元里，越来越多份额会转向带大模型能力的方案，这才是GenAI+教育增长的底盘。

对学校来说，AI采购不再是“尝鲜”，而是信息化预算的必选项，这解释了为什么中小学标书中GenAI占比能冲到46%（见4.3）。政策把“人工智能全学段教育”写进规划，采购就从小心试点变成硬任务。换句话说，教育AI市场的“大”，是建立在政策驱动和预算迁移上的“稳”，不是消费互联网那种爆发式增长，机构做预期管理时要分清这两种曲线。

4.2 玩家分层：四层的优劣势与生死变量

把市场里的玩家分成四层，比列一份公司名录更有用。每一层的优劣势和决定生死的变量都不一样。

大厂（字节豆包爱学、阿里千问智学、腾讯混元、百度等）。优势是算力、资金、流量和工程能力，做通用模型是主场。它们最擅长的是把通用能力产品化，豆包爱学、千问智学（阿里，由“夸克学习”升级、接入千问学习大模型）都直接用自家大模型做教育垂直场景。缺的是教育Know-how：高考命题规律、教材版本差异、家校沟通话术，这些隐性知识大厂没有，得靠收购或合作补。生死变量在于，能否从“提供模型”走到“理解教学”，否则永远停留在流量入口，赚不到教育的深钱。

教育公司转型（学而思、网易有道、粉笔、中公、华图、豆神、学大等）。优势是现成的教研数据、用户和品牌信任，这是大厂短期内买不到的护城河。包袱也重：旧业务还在贡献主要收入，组织里有一大批习惯于旧模式的老师和销售，转型容易脚踩两只船。生死变量是“自研还是接入”以及“数据能不能变成模型壁垒”。粉笔用十年题库和数千万用户行为数据训垂类模型，学大自研“星图”大模型接入DeepSeek，都在赌数据护城河，但多数传统机构对AI研发团队的建设才刚起步。

AI原生创业公司（松鼠AI、各类批改与测评初创等）。优势是轻、专注、模型原生，不被存量业务绑架，敢把产品完全围绕模型设计。松鼠AI用超200亿条学习行为数据和错因体系做自适应，就是典型。活下来的关键是“细分场景的深度加数据飞轮”，在一个窄场景里比大厂做得好、积累专属数据、形成正循环。最大风险是巨头的降维：一旦大厂把同类能力免费化或低价化，初创的付费空间会被快速压缩，所以必须跑得比大厂快、扎得比大厂深。

硬件厂商（科大讯飞、希沃等）。优势是有物理入口和更高毛利，学习机、智能黑板天然绑定使用场景。天花板也明显：硬件换新周期长，卖出去之后“靠内容和服务续费”才是真战场，而这要回到软件能力。2025年教育智能硬件的GenAI渗透率已达50%至60%（来源：艾瑞咨询，2026年），说明硬件已是AI教育最成熟的载体之一，但硬件公司必须证明自己也是“软件和内容公司”，否则只是帮芯片和模型厂商卖入口。

一句话总结四层：大厂有技术缺教研，转型公司有教研缺速度，原生公司有机会缺规模，硬件公司有入口缺内容。各自的生死，都卡在自己最短的那块板上。而2025年之后，四层之间的边界正在模

糊：大厂下场做教育硬件，硬件厂自研大模型，转型公司接通用模型做应用，原生公司争取被集成。混战之下，最终比的不是“属于哪一层”，而是“有没有把某一层的短板补上”。

4.3 四条商业化路径：对比

教育AI的付费方不同，商业模式天差地别。下面这张表把四条主路径放在一起比，凡是没有可靠来源的数字，只写定性。

路径	付费方	决策链长度	客单价量级	毛利特征	获客方式	主要风险
To C 订阅	家长/学生	短（个人决策）	中（月费至数千元/年）	中高，但获客成本高	流量投放、应用商店、口碑	获客贵、续费与效果强绑定
To B 校内	学校/机构	中（教研室到校长）	中高（数万至百万级）	中，定制程度高	直销、渠道、标杆校	决策慢、部署重、回款长
To G 区域	教育局/财政	长（多部门审批）	高（百万至千万级）	项目制偏低，回款慢	招投标、区域集采	预算波动、合规与审计严
硬件	家长/学校	短至中	高（千元至数千元/台）	较高（硬件+内容）	零售、电商、渠道	换新周期长、内容续费难

需要特别点出的是To G的采购画像。艾瑞对GenAI相关标书样本的拆解显示，明确涉及GenAI的项目占比，普通高校27%、职业院校35%、中小学46%（来源：艾瑞咨询，2026年）。中小学反而最高，这是个反直觉的结论：通常认为高等教育才用得起AI，但真实采购里，中小学因为政策驱动（人工智能全学段教育）和示范校建设，反而冲得最猛。预算区间上，高校集中在100万至400万元，双高职400万至千万级、普通高职150万至400万元、中职及偏远院校低于150万元（来源：艾瑞咨询，2026年）。

这张表背后还有一条获客规律：To C靠流量，贵且留存看效果；To B靠标杆校和直销，慢但黏；To G靠关系和合规，一单吃一年但波动大；硬件靠渠道，一次买卖要靠内容续费补。没有哪条路轻松。对中小机构，最务实的是先选一条匹配自己资源的路径做深，而不是四条都碰。区域上，一线市场已被大厂和头部公司占住，下沉到二线示范校、三线补强市场，反而可能是中小玩家的空间。对供应商，这个结构意味着中小学是当下最热的买单方，但决策链条短不代表门槛低。示范校项目高度依赖政策节奏和区域关系，且单校预算有限（多低于150万元），要靠“区域集采”才能上量。从获客成本看，To C最贵也最卷，流量红利见顶后单客成本逐年上升；To G看似单大，但销售周期以年计，且高度依赖政策窗口。这也是市场上出现“先免费后收费”“硬件亏本内容补”等变形打法的原因，本质都是在用一条路径的亏损换另一条路径的留存，对现金流是持续考验。

4.4 单位经济与获客难题：用一个真实样本解剖

这一节是全书最缺资料也最值钱的地方。我们用科大讯飞的2025年报做解剖，因为它是行业里少数披露足够细的教育业务数据的公司。

讯飞2025年营收271.05亿元（同比+16.12%），归母净利润8.39亿元（+49.85%），但扣非净利润只有2.64亿元（来源：科大讯飞2025年报，2026年4月）。智慧教育业务收入89.67亿元（+24.04%），占总营收33.08%，是第一大业务。单看教育产品和服务毛利率55.11%，开放平台毛利率19.36%，综合毛利率42.64%，教育To B/To G业务的毛利显著高于平台业务（来源：科大讯飞2025年报，2026年4月）。

把这几组数放在一起，矛盾就出来了：教育业务毛利率55%看着很漂亮，但公司整体扣非净利率只有约1%，2026年一季度营收52.74亿元（+13.23%）却归母净利润亏损1.70亿元，只是B/C端增速有26.27%（来源：科大讯飞2025年报及2026一季报，2026年4月）。原因在投入端：2025年研发投入53.64亿元，占营收19.79%，研发人员10040人、占59.70%（来源：科大讯飞2025年报，2026年4月）。

所以教育AI的真实难题不是毛利，是研发投入和回款周期。模型要持续训、知识库要持续建、信创适配要持续做，这些都是重投入；而G端项目回款慢、B端采购周期长，现金回正需要时间。讯飞的现金流32.08亿元创历史新高、销售回款274亿元超过收入，说明账面是健康的，但利润被研发吃掉了（来源：科大讯飞2025年报，2026年4月）。

还有一个结构信号：讯飞客户结构里G端占比从39%降到36%，B端加C端合计增长20%，战略业务营收占比从56%升到60%（来源：科大讯飞2025年报，2026年4月）。连行业龙头都在从“卖项目”往“卖产品/订阅”迁移，这印证了我们第二章的判断：垂类模型加RAG加人在回路，才是教育的主路径。学校采购的三大共同特征也与此吻合：一是LLM加RAG加知识图谱（直指幻觉痛点），二是交互Agent化，三是强制私有化部署与国产信创适配（来源：艾瑞咨询，2026年）。

另一条佐证是出海：讯飞2025年出海营收同比增长275%，说明教育AI能力在东南亚、中东等市场有输出空间（来源：科大讯飞2025年报，2026年4月）。成本压力也不是讯飞一家的问题，粉笔2025年因AI投入，研发投入同比增长10.6%、云存储服务费增长68.4%（来源：公开报道，2026年），说明“模型调用加数据存储”在规模化后会变成持续的成本项，不是一次性投入。

把账算清楚后，一个判断浮现：教育AI的胜负不在“谁的模型参数多”，而在“谁能把研发投入转化成可持续的产品化收入和可控的回款周期”。烧钱换规模、靠项目制输血的模式，在合规与预算收紧的周期里最脆弱。

讯飞把“战略业务”（AI相关）占比从56%提到60%，说明它主动把资源往高毛利、强现金流的C/B端收，对G端项目更审慎。这种“退一步、稳现金”的打法，在行业里会越来越像主流：当政策和预算都不确定的时候，活下来比冲规模重要。对不同体量的玩家，启示不同。头部有本钱扛研发、做平台；中小机构别学讯飞重投入自研，更现实的路是“采购模型加自建知识库加做好服务”，把投入集中在自己最懂的教学场景上。教育的账，最终是“教研深度乘技术杠杆”，光有杠杆没有教研，撬不动。讯飞的例子也提醒，“AI+教育”概念再热，资本市场最终看的是扣非利润和现金流，不是营收增速。2025年讯飞营收增16%但扣非仅2.64亿元，这个落差所有讲故事的公司都要正视。

对机构的启示很直接：别只盯着毛利看教育AI的账。真正要算的是，你的研发投入能不能换来可持续的数据壁垒和产品化收入，以及回款周期是否在现金流承受范围内。烧钱买规模、靠项目制输血的模式，在合规和预算收紧的周期里会很难受。

图 6 玩家分层与代表企业（四类玩家优劣势对比）

层级	代表玩家	核心优势	核心短板 / 生死变量
大厂	字节豆包爱学、阿里千问智学、腾讯混元、百度	算力·资金·流量·工程能力	缺教育Know-how；能否从"提供模型"走到"理解教学"
教育公司转型	学而思、网易有道、粉笔、中公、华图、学大	教研数据·用户·品牌信任	旧业务包袱；自研还是接入；数据能否变壁垒
AI原生创业	松鼠AI等	轻·专注·模型原生；细分深度+数据飞轮	巨头降维免费化；缺规模
硬件厂商	科大讯飞、希沃	物理入口·较高毛利	换新周期长；须证明也是软件/内容公司

四层边界正在模糊：大厂下场做硬件、硬件厂自研大模型、转型公司接通用模型、原生公司争取被集成（见第四章4.2）。

图 7 四条商业化路径单位经济风险画像

路径	获客难度	毛利水平	回款周期	决策链	客单价
To C 订阅	高	中高	短	短	中
To B 校内	中	中	中长	中	中高
To G 区域	中高	偏低	长	长	高
硬件	中	较高	短-中	短-中	高

红=挑战大 黄=中性 绿=相对有利；详见第四章4.3 4.4。

第五章 政策与合规红线

前三章讲的是技术能做什么、用在哪、市场怎么分。这一章把问题反过来问：在教育这个面向未成年人的高频场景里，哪些事是“不能做”或“做之前必须先过关”的。

教育行业的一个特殊之处，是服务对象里有大量未成年人，而未成年人保护是中国网络空间治理里监管最密、处罚最严的一条线。对教育AI来说，合规不是上线之后的“补丁”，而是产品设计的“地基”。一个面向中小学生的AI学伴，从第一天起就要同时满足大模型备案、未成年人保护、数据安全三套要求，任何一条缺位，都可能不是“被约谈”，而是“下架”。

5.1 政策时间线：2024年至2026年的关键节点

把近两年与教育AI直接相关的政策按时间排开，能清楚看到一条从“鼓励探索”到“规范落地”的线。

2023年8月15日，《生成式人工智能服务管理暂行办法》施行，这是生成式AI服务的上位法，明确了训练数据合法来源、防范未成年人沉迷、内容标识、算法备案与安全评估等基本要求（来源：国家网信办等七部门，2023年）。

2024年1月1日，《未成年人网络保护条例》正式施行，其中专门规定以未成年人为服务对象的在线教育网络产品和服务，须按不同年龄阶段身心特点和认知能力提供相应产品（来源：国务院令 第766号，2024年）。

2025年是政策密度最高的一年。4月，教育部等九部门印发《关于加快推进教育数字化的意见》（教办〔2025〕3号），首次把算法与大模型备案机制写进教育政策；5月，教育部基础教育教学指导委员会发布《中小生成式人工智能使用指南（2025年版）》和《中小学人工智能通识教育指南（2025年版）》，第一次对“学生能不能用、怎么用生成式AI”给出分学段规范；6月，在中央网信办指导下，中国网络空间安全协会会同60家单位发布《向未成年人提供生成式人工智能服务安全指引》团体标准，同月国务院印发《教育发展“十五五”规划》，明确“推进人工智能全学段教育”（来源：教育部、中央网信办等，2025年）。

2025年9月1日，四部门《人工智能生成合成内容标识办法》施行，要求AI生成的文本、图片、音视频等必须添加显式或隐式标识（来源：国家网信办、工信部、公安部、广电总局，2025年）。2026年3月1日，八部门《可能影响未成年人身心健康的网络信息分类办法》施行，明确算法推荐、生成式AI等“不得向未成年人推送可能影响其身心健康的网络信息”（来源：国家网信办等八部门，2026年）。

这条时间线的含义很直接：到2026年，教育AI已经不是在“等政策”，而是“政策已经铺好，看你怎么合规地跑”。对机构来说，另一个含义是：合规能力正从“成本项”变成“竞争力”。当政策

把备案、白名单、私有化都变成硬要求，能更快合规的机构，反而能更早拿到校内和G端的入场券。

5.2 三道红线：未成年人保护、数据安全、算法备案

把所有要求收敛一下，面向教育机构，最要紧的是三道红线。

红线一，未成年人保护。核心不是“不能用AI”，而是“怎么用才不伤孩子”。《未成年人网络保护条例》要求在线教育产品按年龄阶段适配（来源：国务院令 第766号，2024年）；2026年3月施行的《分类办法》进一步禁止向未成年人推送可能影响身心健康的内容（来源：国家网信办等八部门，2026年）；教育部《使用指南》则做了更细的分学段规定：小学阶段学生在教师、家长帮助下使用开放式内容生成功能，初中可适度探索生成内容的逻辑性分析，高中结合技术原理开展探究性学习，同时明确禁止学生直接复制AI生成内容作为作业或考试答案，杜绝“代劳式”使用（来源：教育部基础教育教学指导委员会，2025年）。对产品而言，未成年人模式、使用时长限制、内容分级、防沉迷都不再是“可选项”。分学段不是形式主义，而是承认低龄孩子的判断力还不足以甄别AI错误，需要成人兜底。

红线二，数据安全。教育数据是高度敏感的个人数据。个保法和《暂行办法》都要求训练数据来源、不收集非必要个人信息、不非法留存用户输入（来源：国家网信办等七部门，2023年）。落到学校场景，《使用指南》明确严禁师生输入考试试题、个人身份信息敏感数据，并要求学校建立生成式AI工具“白名单”制度，只放行数据安全合规的工具（来源：教育部基础教育教学指导委员会，2025年）。结合学校采购“强制私有化部署与国产信创适配”的特征（来源：艾瑞咨询，2026年），教育AI的数据底线是：学生数据不出校、不上通用公有云、敏感信息不进模型训练。

红线三，算法备案。提供具有舆论属性或社会动员能力的生成式AI服务，须开展安全评估，并按《互联网信息服务算法推荐管理规定》履行算法备案手续（来源：国家网信办等七部门，2023年）；提供AI生成合成内容时，还要按《标识办法》提交标识相关材料（来源：国家网信办等四部门，2025年）。截至2025年11月1日，全国累计已有611款生成式AI服务完成备案、306款应用或功能完成登记（来源：国家网信办，2025年11月）。对机构来说，自研或深度定制模型上线前，备案是绕不开的前置动作；直接调用已备案模型的，也要确认模型备案号并在产品页公示。

把三道红线放在一起看，它们不是并列的清单，而是一道递进的合规闸门：算法备案决定“能不能上线”，数据安全决定“能不能碰学生数据”，未成年人保护决定“能不能服务孩子”。备案通道已经跑通，机构不必把它当成无法逾越的障碍，但不能绕过它。

5.3 内容合规：价值导向与科学性

红线之外，是两条贯穿日常的内容要求。

一是价值导向。《暂行办法》要求坚持社会主义核心价值观，不得生成违法有害内容，在算法设计、训练数据、模型优化中防止歧视（来源：国家网信办等七部门，2023年）。对教育产品，这条线比普通应用更紧，因为受众是价值观正在形成的未成年人。一个讲历史、讲社会的AI答疑，输出里夹带偏见或错误史观，性质不只是“答错”，还可能是“内容违规”。

二是科学性，也就是事实准确。前面第二章讲过幻觉的三类表现，但在这里要换个角度说：在教育场景，事实错误不只是体验问题，更是合规风险。当AI生成的内容被当作学习材料分发给学生，错误知识一旦传播，就触及内容安全。加上《标识办法》要求自2025年9月1日起AI生成内容必须标识（来源：国家网信办等四部门，2025年），“这是AI写的”必须让用户知道，否则同样违规。教育部《使用指南》把“白名单”和“敏感数据禁输入”并列，正是因为这类错误发生在校园、面向未成年人，性质不同于面向成人的聊天机器人。

这两条合起来，指向同一个工程要求：教育AI的输出，既要在价值上“对”，也要在事实上“准”，且要“说清楚这是AI生成的”。价值偏差靠审核机制，事实错误靠RAG加人审，身份混淆靠显式标识，三者缺一不可。

5.4 机构最容易踩的四个坑

把红线落到实操，我们观察到的机构高频踩坑有四个。

坑一，拿通用工具直接进校。不少机构把面向成人的通用大模型或聊天机器人，原样放到校园场景，没做未成年人适配、没进白名单、没设使用时长。这类产品一旦被认定为“向未成年人提供未合规的服务”，风险极高。正确做法是：进校前先过白名单审查、做年龄分层和防沉迷。

坑二，默许甚至鼓励“代劳式”使用。一些机构把“AI一键生成作文、作业”当作卖点，直接踩了《使用指南》的红线。合规的边界是：AI是辅助，不能替代学生的独立思考与产出，作业和考试答案不能由AI直接代写。

坑三，把学生数据喂给通用公有云模型。为了省事，把本校试题、学生身份信息输入未私有化部署的通用模型，既违反“敏感数据禁输入”，也违反数据本地化与信创要求。正确做法是私有化或专属知识库部署，学生数据本地化处理。

坑四，AI生成内容不标识、不人审直接出分。按《标识办法》，AI生成内容必须标识身份；按本书的判断（第二章），凡是判分环节必须人审。学而思学习机曾出现“把对的判错”“数学题给英语答案”（来源：新浪财经、红星资本局，2024至2025年，见本书第二章），若这类错判未经人审直接记入成绩，就是一起教学事故。跳过标识和人审这两步，短期省事，长期是确定性违规。

一句话收尾：教育AI的合规，不是“上线后补课”，而是“设计时就埋进地基”。谁先把三道红线和四类坑变成产品默认逻辑，谁才可能在监管常态化之后活下来。监管常态化之后，合规会从“应付检查”变成“日常运营的一部分”，越早把它产品化，后续合规成本越低。

图 8 教育AI合规红线与责任主体

红线/要求	主要法规	核心要求	责任主体	违规后果
未成年人保护	《未成年人网络保护条例》(国务院令766号, 2024)+《分类办法》(2026.3)+教育部《使用指南》(2025)	按年龄适配、未成年人模式、防沉迷、禁推送有害信息、禁代劳式使用	产品(年龄分层/防沉迷)+学校(分学段管理)	下架/处罚
数据安全	个保法+《暂行办法》+教育部《使用指南》	训练数据合法来源、敏感数据禁输入、学生数据不出校/不进公有云、白名单	供应商(私有化/信创)+学校(白名单)	违规采集/泄露处罚
算法备案	《算法推荐管理规定》+《暂行办法》+《标识办法》	安全评估+算法备案(截至2025.11.1共611款备案/306款登记)+AI内容标识	供应商(备案/公示备案号)	未备案不得上线
内容·价值导向	《暂行办法》	坚持社会主义核心价值观、防歧视	产品(审核机制)	内容违规
内容·事实准确	第二章判断+《标识办法》	RAG+人审、防幻觉	产品(人审)+学校	错误知识传播=内容安全事件

红线/要求	主要法规	核心要求	责任主体	违规后果
内容·身份标识	《标识办法》(2025.9.1施行)	AI生成内容显式/隐式标识	产品(标识)	未标识违规

三道红线是递进闸门：备案决定"能否上线"、数据安全决定"能否碰学生数据"、未成年人保护决定"能否服务孩子"（见第五章）。

第六章 落地方法论：机构到底该怎么做

前面五章回答了“能做什么、用在哪、市场怎么分、红线在哪”。这一章回到最务实的问题：一个学校、一家教育公司、一个区域教育部门，具体该怎么把AI落进日常？我们不给万能公式，但给一套可操作的判断框架。

6.1 能力评估自查清单

- 动手之前，先用这张清单给自己打分。每一项“否”，都是后续要补的短板。
- 一是数据底座：有没有结构化的校本知识库（教材、课标、题库、教研沉淀）？没有它，模型就是“普遍正确但对你没用”。
- 二是技术选型：是否采用“垂类模型 + RAG + 人在回路”的主路径（第二章结论），而不是裸接通用大模型？
- 三是合规前置：产品是否通过算法备案或登记、是否进校白名单、是否私有化部署、是否设未成年人模式？
- 四是场景边界：是否明确哪些环节AI只做“初筛和标注”、哪些必须由人终审（如判分）？
- 五是组织准备：教师数字素养、AI使用培训、家校沟通机制是否就位？
- 六是效果验证：有没有第三方或对照组的客观效果证据，而不是只靠企业自陈？
- 清单之外还有一条隐性要求：选供应商时确认其具备信创适配与私有化交付能力，否则即便模型好，也过不了学校的采购审查（来源：艾瑞咨询，2026年）。

这张清单的本质，是逼机构分清“我买了AI”和“我真的能用AI”。前者是采购，后者是能力。我们建议机构逐项自评，任何一项打“否”都先补齐再谈规模化。尤其要警惕一种错觉：以为“模型接上了”就等于“能力具备了”。本书第四章解剖的讯飞样本显示，连行业龙头都要把营收的近20%投在研发上，中小机构更该把有限资源压在“知识库”和“组织准备”这两项软基建上，而不是追模型参数（来源：科大讯飞2025年报，2026年4月）。这份清单偏向供给侧能力，机构还要同时评估需求侧：学生和家长为什么用、愿不愿意为它改变习惯，这部分在本书第三章的用户行为调研里已有铺垫。一个可操作的起点，是先做一个最小可行的小场景：选一个班级、一门学科、一个场景，跑通“提问、检索、人审、交付”的流程，再用它去争取校内其他学科的支持。清单可以每半年复评一次，因为教育AI的能力边界和政策要求都在快速变化。

6.2 场景优先级：价值与难度四象限

- 不是所有场景都该同时做。我们建议用“教育价值”和“落地难度”两个维度排序，优先从“高价值、低难度”的象限切入。
- 高价值、低难度（优先做）：智能批改与作业系统、教师备课工具。任务边界清晰、易量化、风险可控，是验证团队能力的稳妥起点，也能最快让教师感受到减负（来源：本书第三章、第四章判断）。先做这一步的好处是“以小胜立信”：教师看到批改时间真的少了，才愿意在学伴、评价等更难

的场景里继续试。

高价值、高难度（重点攻坚）：AI学伴、个性化辅导。价值最大但“AI独立教学”仍处早期，适合作为长期方向而非第一步（来源：本书第三章判断）。这类场景的难点不在技术峰值，而在“持续跟踪一个学生三个月的起伏”，目前只有绑在硬件或强账号体系里的产品摸到门槛。

低价值、低难度（谨慎投入）：部分营销、运营类AI应用，能做但不该挤占核心资源。前者常见如AI招生文案，顺手做即可。

低价值、高难度（暂不做）：泛化的“全能AI老师”，投入巨大且效果未证，是多数机构的陷阱，建议直接放弃。把资源压在这种“看起来很美”的方向上，是中小机构最常见的战略失误。

学校采购的三个共同特征值得对照：LLM + RAG + 知识图谱、交互Agent化、强制私有化与信创适配（来源：艾瑞咨询，2026年）。机构的场景设计，最好与之对齐，否则进不了白名单、过不了审查。

还有一个常被忽略的战略变量：孩子接触AI的首要途径是通用AI助手，占比74.8%，而非某个教育垂类App（来源：艾瑞咨询，2026年，见本书第三章）。这意味着机构自建的学伴App未必是孩子第一个打开的入口，更现实的策略，要么把教育能力嵌进通用助手，要么靠学习机或强账号体系把学生留在自己的场景里。排场景优先级时，要把“入口归属”也算进去。

图 9 教育AI场景优先级：教育价值 × 落地难度

纵轴：教育价值（低 → 高） 横轴：落地难度（低 → 高）	
高价值 · 低难度（优先做） 智能批改与作业系统（边界清晰、易量化、风险可控） 教师备课工具（减负直接、最快立信）	高价值 · 高难度（重点攻坚） AI学伴与个性化辅导（价值最大，“AI独立教学”仍早期，宜作长期方向）
低价值 · 低难度（谨慎投入） 部分营销/运营类AI（顺手做即可，不挤占核心资源）	低价值 · 高难度（暂不做） 泛化“全能AI老师”（投入巨大、效果未证，多数机构陷阱）

优先级判断见第六章6.2；建议从“高价值 · 低难度”象限切入，以小胜立信。

6.3 自研、采购还是共建：决策框架

路径选择取决于机构自己的资源。三种路径不是三选一，而是常按阶段组合：先用采购跑通场景，再用共建补合规与数据，最后才谈自研。

自研，适合有海量专属数据、强算力与算法团队、且业务高度依赖模型差异化的头部玩家（如学而思、网易有道、讯飞）。代价是重投入、长周期，讯飞2025年研发投入53.64亿元、占营收19.79%就是例证（来源：科大讯飞2025年报，2026年4月）。但自研的隐性风险在组织：旧业务团队与AI团队目标冲突、数据孤岛、研发投入短期看不到回报时的内部质疑。多数传统教育机构卡在“想自研但养不起团队”。

采购，适合绝大多数中小机构。直接调用已备案的通用或垂类模型，把投入集中在自建知识库和教学服务上。这是性价比最高、也最可控的起点，尤其契合“模型越来越便宜，但把模型变成好老师依然昂贵”的现实（来源：本书第二章判断）。随着开源模型拉低调用成本，采购路线的经济性还在改善。采购也有光谱：纯API调用最轻但数据出校风险高，私有化部署贵但合规，前者适合C端轻场景、后者适合校内与G端。选的时候把“数据在哪里”作为第一判断。

共建，适合学校与区域教育部门。学校出教学场景和私有数据，厂商出技术与合规能力，联合打磨。对G端和示范校，共建比纯采购更稳，也比自研更现实。天立教育与腾讯共建“学科大脑”即是一例：学校出场景与数据，腾讯出混元大模型，双方合起来覆盖教研教学全场景（来源：封面新闻，2025年12月，见本书第一章）。共建也有前提：学校要舍得开放真实教学数据，厂商要守得住合规，否则容易变成“演示很热闹、平时没人用”的摆设。

路径选择还要看付费方：To C靠流量、To B靠标杆、To G靠关系（见本书第四章），自研和共建的回报周期都比采购长，现金流不厚的机构慎选。无论选哪条路，本书第二章的判断都成立：模型越来越便宜，但把模型变成好老师依然昂贵，差异最终落在教研深度上。一句话：别学龙头重投入自研，也别指望买个模型就万事大吉；多数机构的现实答案，是“采购模型 + 自建知识库 + 做好服务”（来源：本书第四章讯飞启示）。

6.4 从试点到规模化的推进节奏

很多项目死在“既要又要”：想一次覆盖所有学段、所有学科、所有角色。我们建议克制，用三步走把野心拆小，每一步都有明确的“过关条件”再进下一步。每一步都建议配一名“AI推进负责人”，否则试点很容易在教师轮换、学期结束时不了了之。

第一步，单点试点（1至3个月）：选一个高价值低难度场景（如批改或备课），在小范围教师中跑通，验证工具可靠、教师愿用、合规无碍。过关条件：教师周活过半、无合规投诉、判分零事故。不过关不扩张。

第二步，校内推广（3至6个月）：沉淀校本知识库，形成稳定的人审流程，做教师培训和家校告知，把试点经验固化成可复制的规范。过关条件：知识库覆盖本校主要学科、人审规范成文、家长知情同意。

第三步，规模化与区域联动（6个月以上）：从单校到多校、从单场景到多场景，介入区域集采或G端项目，同时建立持续的效果评估机制。过关条件：跨区域效果可复现、有独立评估结论。效果评估不必复杂，可以从“教师周节省时长”“学生错题复发率”“家长满意度”三个指标起步，用前后对照说话。

节奏的关键词是“慢试点、快复制”：前期宁可小范围把坑踩完，也不要一上来就全校铺开，否则一个判分事故就可能让整个项目退回原点。反例比比皆是：某校一口气给全校装了AI学习终端，教师没培训、知识库是空的，两个月后使用率跌到个位数，项目不了了之。问题不在工具，在跳过了前两步。

6.5 五种常见失败模式与规避

这些失败模式有个共同前提：机构把“上线”当成终点，而教育AI的终点其实是“持续可用”。最后列五类我们见过的高频失败，给后来者提个醒。

一是“重模型、轻知识库”：花大钱接最强模型，却没建校本知识库，结果“普遍正确但对你没用”。某区域采购了强模型但没接校本题库，学生问“本校月考考点”模型瞎编，教师还得重答。知识库不是一次性工程，需要随教材和考点持续更新，否则半年后模型又回到“普遍正确但对你没用”。规避：知识库优先于模型升级。

二是“重采购、轻培训”：买了系统教师不会用，使用率个位数。典型画面：系统买来，全校大会演示一次，之后锁在机房，期末盘点使用率3%。规避：把教师培训写进预算，而非上线后才补。

三是“重功能、轻合规”：为赶时髦上线未备案、未白名单、未私有化的产品，随时可能被叫停。合规缺位的产品，一旦被监管认定违规，轻则整改、重则下架，前期投入归零。规避：合规前置到选型阶段。

四是“重演示、轻人审”：Demo惊艳，但判分、结论直接出给学生，错一次就失信。学而思学习机曾出现“把对的判错”“数学题给英语答案”（来源：新浪财经、红星资本局，2024至2025年，见本书第二章），若这类错判未经人审直接进成绩，就是事故。规避：关键节点强制人审。

五是“重采购、轻评估”：上线后没有效果追踪，半年后说不清AI到底有没有用。半年后校长问“这钱花得值不值”，无人能答，续费自然暂停。规避：从第一天起建效果基线和对标。

这五种失败，根子都在同一个地方：把AI当成“买来就生效的魔法”，而不是“需要人、数据、流程一起搭的基建”。本白皮书第七章会进一步讨论AI在教育里的效果边界与冷思考，第八章给出面向不同机构的差异化行动建议。

第七章 风险的冷思考

注 生成式AI给教育带来的，不只是一串增长曲线，更是一组必须被严肃对待的结构性风险。这一章不唱反调，而是把风险摆到桌面上——因为只有看清风险，落地才有章法。

7.1 一个并不讨喜的"40/60分割"

在和数十位校长、产品负责人聊过之后，我们发现一个反复出现的判断：生成式AI对教育的价值，粗略可以拆成两块。

约 40% 来自"效率提升"——备课、出题、批改、答疑的自动化，这部分风险相对可控，本质是用机器替代重复的脑力劳动，边界清晰、收益可测。约 60% 来自"教与学关系的重构"——个性化学习路径的生成、师生互动方式的改变、评价体系从标准化走向过程性。这部分才是真正的价值高地，也恰恰是风险最集中的地带。

一个常见的误判，是把风险想成只存在于表层工具（"AI写错了答案"），而忽略了深层变革区的不确定性（"当AI替学生想好了，学生还学不学思考"）。本章的风险清单，按这个"40/60"的逻辑展开：表层的、可控的，和深层的、需要制度性应对的。

7.2 六大风险维度

7.2.1 幻觉与事实性错误

这是用户感知最直接的一类风险。在作业辅导、考试备考场景里，学生最依赖AI给出"标准答案"，但大模型的幻觉在数理推导步骤、文史细节、冷门知识点上仍时有发生。我们已经看到案例：某解题产品把一道几何题的错误辅助线讲得头头是道，误导了学生整整一个晚自习。

可落地的缓解手段并不神秘：第一，用 RAG（检索增强生成）把答案锚定在教材与权威题库，而非纯靠模型"回忆"；第二，关键结论强制"溯源"——要求引用教材页码或知识点编号；第三，产品层面对低置信度答案主动给出"我不确定，请核对"的提示，而不是假装笃定。把"可靠性"当成功能来设计，而不是指望模型自己变准。

7.2.2 数据隐私与未成年保护

教育数据是敏感性最高的数据类别之一：人脸、声纹、课堂行为、学情轨迹，几乎覆盖了一个未成年人成长的完整数字画像。合规依据是清晰的——《个人信息保护法》《未成年人保护法》以及教育部对教育数据安全的相关要求都明确，未成年人的个人信息处理必须最小化、必须告知同意、必须严格境内存储。

真实风险在于三个环节：厂商在免费产品中过度采集；数据在未脱敏状态下被用于模型训练；以及跨境传输带来的管辖权空白。行业里已经出现"用学生作文训练自家模型"的争议案例。缓解路径是工程化加制度化的组合：优先私有化部署、数据最小化采集、家长知情同意机制、并在合同里明确"禁用未成年数据训练"。对采购方而言，这应写进招标的技术条款，而不是事后补救。

7.2.3 算法偏见与公平性

大模型的训练语料天然带有偏见，这些偏见会顺着推荐内容、评价打分、资源分配悄悄进入教育场景。更棘手的是，AI本被寄望于"促进教育公平"，但如果算力、数据、优质模型集中在少数发达地

区与头部机构，结果可能是"AI平权"变成"AI分化"——资源越好的学校用得越好，薄弱校连基础算力都拿不到。

缓解不能只靠厂商自觉。区域教育云的统建、开源/国产模型的算力下沉、以及政策层面对薄弱校的定向投入，是把"公平"写进技术架构的必要动作。一个务实的原则是：凡是涉及评价与资源分配的系统，必须做偏见审计并公开审计结论。

7.2.4 学术诚信与认知外包

这是教育工作者最焦虑的一类风险。当学生用AI代写作文、解作业、写实验报告，"思维能力外包"就发生了——表面完成了任务，底层的能力没长出来。高校里围绕"论文AI率检测"的争议，本质是旧评价体系撞上了新生产力。

缓解的关键，不是堵（堵不住，也堵错了方向），而是改。过程性评价取代一次性成果评价；把AI定位为"协作者"而非"代写"——例如要求学生展示"我如何与AI一起完成了这个项目"，而不只是交一份成品；口试、实操、项目制学习重新成为能力甄别的主场。更重要的是开设"AI素养"课，让学生理解工具的边界：什么时候该用，什么时候必须自己想。

7.2.5 教师角色重构与组织阻力

一个需要反复澄清的误解：生成式AI不是来"替代教师"的，它是来"重构师生关系与师师协作"的。教师从知识的唯一传授者，变成学习体验的设计师、AI的编排者、以及学生心智成长的陪伴者。这个角色其实更高阶，也更难。

阻力来自三处：学校的组织惰性（排课、教研、评价指挥棒都没变）、教师自身的AI素养缺口、以及"引入AI=给老师增负"的现实感受。缓解要对症下药：借《新时代基础教育强师计划》等政策红利做教师AI能力培训；建AI教研共同体让好经验流动；最关键的一条——任何AI工具必须以"给教师减负"为第一目标，凡是增加教师事务性负担的产品，都不该进课堂。

7.2.6 厂商锁定与可持续性

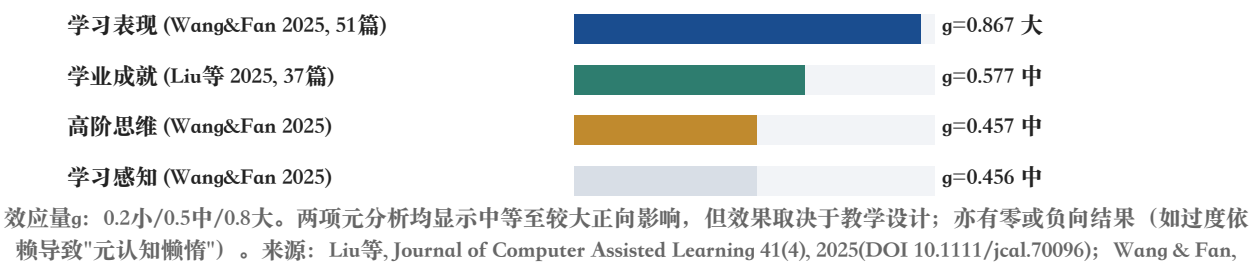
最后一类风险常被忽视，却是商业落地的命门：教育AI的单位经济相当脆弱。以行业标杆科大讯飞为例，其2025年年报显示智慧教育业务毛利率约55.11%，看似不错，但扣非归母净利润仅约2.64亿元，2026年一季度仍然亏损。这个反差说明一件事——教育AI不是"毛利不够高"的问题，而是"研发投入重、G端回款周期长、C端获客贵"的问题。

由此带来的风险是：用烧钱换份额、被通用大模型大厂降维打击、以及用数据孤岛把客户牢牢锁定却无力持续迭代。对采购方和创业者共同的提醒是：看现金流健康度胜过看毛利率；产品应开放接口、与既有教务系统解耦，避免被单一厂商技术绑架。

7.3 证据的另一面：有效性并非空谈

谈完风险，也要避免滑向另一种极端——认为"AI对学习教育没用"。事实上，已有严格的循证研究支持其在受控场景下的正向效果。

图 10 GenAI学习效果：代表性元分析效应量对比 (Hedges g)



2025。

图10汇总了两项近期元分析的核心效应量（effect size，以 Hedges' g 计，数值越大代表效果越强）：Liu、Zuo 与 Lu（2025）发表于 *Journal of Computer Assisted Learning*（41卷4期，DOI 10.1111/jcal.70096）对37项研究的元分析显示，学业成就效应量 $g=0.577$ ；Wang 与 Fan（2025）对51项研究的元分析显示，学习表现 $g=0.867$ 、学习感知 $g=0.456$ 、高阶思维 $g=0.457$ 。

这些数字的意义在于：它把"AI有没有用"从观点拉回到证据层面。但效应量不等于"无条件适用"——它衡量的多是受控干预场景下的平均效果，真实课堂里能否复现，取决于前面六大风险是否被工程化地处理好。风险与证据，是同一枚硬币的两面。

7.4 收束：风险是落地清单，不是退缩理由

把这一章的六大风险翻译成一份可执行的清单，反而最有用：

1. 事实性：所有面向学生的答案是否可溯源、是否做过幻觉压测？
2. 隐私：是否私有化部署、是否最小化采集、是否禁用未成年数据训练？
3. 公平：涉及评价与分配的系统是否做了偏见审计并公开？
4. 诚信：评价体系是否从"成果"转向"过程"，是否教了AI素养？
5. 教师：工具是否以"给教师减负"为第一目标？
6. 可持续：供应商现金流是否健康、接口是否开放、是否被技术绑架？

风险不是不做的理由。风险是"怎么做才对"的施工图。

第八章 趋势展望

注 如果说2026是"拐点之年"，那真正的分水岭并不在2026，而在2027到2029。这一章不预测具体份额，而是给出几条大概率成立的演进主线，以及从业者"现在就能做"的动作。

8.1 主线一：从"工具"到"学伴"与"助教"

第一代教育AI产品，形态是"工具"——一个聊天框、一个批改器、一个搜题器，人找功能。下一阶段会走向"学伴"与"助教"——AI主动理解学生的学习状态，持续陪伴、主动提醒、动态调整节奏。

更值得关注的是 Agent 化：单个大模型变成多个协作的智能体（一个负责讲解、一个负责出题、一个负责复盘、一个负责与家长/教师沟通），它们围绕一个学生协同工作。这比"一个万能对话框"更可靠，也更容易在工程上做责任切分。2026年我们看到的还是演示，2027-2028年会有第一批真正跑通的多智能体教学系统。

8.2 主线二：多模态与端侧模型

今天的AI主要处理文字，下一阶段会原理解图像、语音、视频、手写、实验操作。一个学生拍下自己的草稿，AI就能看懂思路卡在哪；一节实验课的视频，AI能自动生成过程性评价。

与此同时，端侧模型（跑在平板、学校本地服务器上的小模型）会重要性上升。原因很现实：隐私友好（数据不出校）、离线可用（不依赖网络）、成本低（不用每次调用云端大模型按token计费）。端侧+云端的混合架构，会成为学校采购的默认选项，而不是可选项。

8.3 主线三：个性化从"推荐"到"共建"

过去十年的"个性化", 本质是推荐——系统根据成绩推题、推课。生成式AI带来的范式变化是"共建": 学生可以参与生成属于自己的学习路径、用自己的语言重写知识点、把抽象概念变成自己能懂的故事。个性化不再是"系统给我什么", 而是"我和系统一起造什么"。

这对产品设计的含义很深: 界面要从"展示"走向"共创", 给学生足够的编辑权与表达空间, 同时保留教师的把关位。

8.4 主线四：评价体系重构

当"写"和"算"可以被AI低成本完成, 以标准化纸笔考试为核心的人才甄别机制必然松动。过程性评价、能力本位评价 (competency-based)、以及"会用人力智能解决问题"的实操考核, 会成为新的指挥棒。

这对学校的冲击最大, 也最慢。但方向是清楚的: 评价什么, 就教什么; 当评价开始衡量"与AI协作解决问题的能力", 课堂才会真正改变。

8.5 主线五：监管与标准成熟

2026年的一个明显变化是, 监管从"观望"走向"建标准"。数据合规、未成年保护、AI生成内容的标识、以及教育AI的质量评估框架, 会在未来两三年里逐步成型。AI素养会从选修变成必修课——不仅是学生的, 也是教师的、管理者的。

对从业者而言, 标准成熟是好事: 它把"合规"从模糊成本变成清晰边界, 反而降低了创新的不确定性。

8.6 主线六：区域与校际分化加剧

AI会放大先发者的红利——最早把数据、流程、师资跑通的学校, 会和后来者拉开代差。但公共政策会托底公平: 区域教育云、国家中小学智慧教育平台、以及对薄弱校的定向算力与培训投入, 是防止"AI鸿沟"变成"AI断层"的关键。

分化不可避免, 断裂可以预防。这是决策者要盯住的平衡点。

8.7 一张路线图

图 11 2026–2029 教育AI关键变量路线图

2026	规范采购与规模应用元年 政策从"鼓励探索"转向"规范采购+规模应用"; 学校为模型推理单独立项, G端/校端采购成主引擎。
2027	垂类模型+RAG成为标配 幻觉治理成熟, 知识库质量成差异化核心; Agent从演示走向教研教学全流程编排。
2028	多模态与个性化规模化 GenAI+教育产品服务规模预计达8910亿元(艾瑞); 自适应与学情画像进入常态教学。
2029	AI基础设施化定型 AI成教学默认运行环境; 效果评估与合规成日常运营; 单位经济改善、现金流为正者胜出。

为基于政策线与市场线的前瞻判断, 非确定预测; 具体节奏受政策窗口与预算影响。

图11把上述主线收敛为四个阶段: 2026年"试点验证"——单点场景跑通、口碑形成; 2027年"规模部署"——区域教育云与多智能体系统进入采购清单; 2028年"深度融合"——AI嵌进教研、评价、管理的日常流程; 2029年"范式确立"——"与AI协作"成为教与学的新常态, 评价体系完成重构。

8.8 给从业者的三条"现在就做"

- 1. 把数据底座搭对：现在就做私有化部署与数据治理，比两年后补课便宜十倍。数据架构错了，上面什么都跑不稳。
 - 2. 重做教师工作流：选"给教师减负"的切入点（出题、批改、学情汇总），先让老师尝到甜头，再谈深层变革。凡是增负的AI，活不过一学期。
 - 3. 建立AI素养与合规双轨：现在就把AI素养课、数据合规审计、偏见审计写进制度。标准成熟之前主动合规，标准成熟之后你就是标杆而非整改对象。
- 拐点已经到来。真正的分水岭，属于那些今天就开始把风险翻译成施工图的人。

附录

一、数据来源清单

本白皮书核心数据均来自公开可查的第三方报告、上市公司年报与同行评议研究。以下按文中出现顺序列出，供读者核验。

[1] 艾瑞咨询.《2025年中国AI+教育行业发展研究报告》. 艾瑞咨询 (iResearch), 2025. (报告中教育信息化经费、C端个性化学习、GenAI相关技术能力价值等规模与CAGR数据均引自该报告公开摘要。)

[2] 中国互联网络信息中心 (CNNIC).《第55次中国互联网络发展状况统计报告》. 2025. (网民规模、生成式AI用户规模等宏观数据来源。)

[3] 科大讯飞股份有限公司.《2025年年度报告》. 2025. (智慧教育业务毛利率、扣非归母净利润等财务数据。)

[4] 科大讯飞股份有限公司.《2026年第一季度报告》. 2026. (一季度盈亏情况。)

[5] Liu, Y., Zuo, M., & Lu, J. Generative AI in education: A meta-analysis of its effect on academic achievement. *Journal of Computer Assisted Learning*, 2025, 41(4). DOI: 10.1111/jcal.70096. (37项研究元分析, 学业成就效应量 $g=0.577$ 。)

[6] Wang, X., & Fan, X. The impact of generative artificial intelligence on student learning: A meta-analysis. 2025. (51项研究元分析, 学习表现 $g=0.867$ 、学习感知 $g=0.456$ 、高阶思维 $g=0.457$ 。)

[7] 中华人民共和国教育部.《教育信息化2.0行动计划》. 2018. (教育信息化顶层政策文件。)

[8] 中华人民共和国教育部等.《新时代基础教育强师计划》. 2022. (教师能力建设相关政策, 文中教师AI素养培训依据。)

[9] 中华人民共和国教育部.《义务教育课程方案(2022年版)》及系列课程标准. 2022. (过程性评价、素养本位导向依据。)

[10] 中华人民共和国中央人民政府.《中华人民共和国个人信息保护法》. 2021年施行. (未成年人与敏感个人信息处理合规依据。)

[11] 中华人民共和国中央人民政府.《中华人民共和国未成年人保护法》(2020年修订). 2021年施行. (未成年人数据保护依据。)

注 说明：部分市场规模与占比数据为行业研究机构公开摘要口径，文中已注明来源年份；个别细分口径（如校端/C端拆分）若公开来源未明确，本书采用更保守的"技术能力价值"口径并标注，避免对未核实数字做精确拆分。

二、术语表

生成式AI (GenAI, Generative AI) : 能基于提示生成文本、图像、音频、视频等内容的AI系统, 代表如大语言模型, 区别于只做分类或预测的传统AI。

大语言模型 (LLM, Large Language Model) : 以海量文本训练、能理解和生成自然语言的深度学习模型, 是当下生成式AI的核心底座。

检索增强生成 (RAG, Retrieval-Augmented Generation) : 先从一个权威知识库检索相关内容、再让模型据此作答的技术, 用来降低幻觉、保证答案可溯源, 是教育AI可靠性的关键工程手段。

智能体 (Agent) : 能自主感知、规划、调用工具并完成多步任务的AI系统; 在教育场景中指能讲解、出题、复盘、沟通多角色协作的AI。

多模态 (Multimodal) : AI同时理解和生成文字、图像、语音、视频等多种形态信息的能力。

微调 / 指令微调 (Fine-tuning / Instruction Tuning) : 在通用模型基础上, 用特定领域数据进一步训练, 使其更贴合教育场景的表达与知识。

幻觉 (Hallucination) : 模型生成看似合理但与事实不符的内容, 是教育AI最需要防控的问题之一。

知识图谱 (Knowledge Graph) : 用"实体—关系"结构组织的领域知识网络, 常与RAG配合, 给AI提供结构化的学科逻辑。

私有化部署 (On-premise Deployment) : 把模型和系统部署在学校或机构自有服务器内, 数据不出域, 满足隐私与合规要求。

端侧模型 (On-device Model) : 运行在平板、手机或本地服务器上的轻量模型, 隐私友好、可离线、成本低, 与云端大模型形成混合架构。

效应量 (Effect Size, 常用 Hedges' g) : 衡量干预效果强度的统计量, 0.2为小、0.5为中、0.8为大; 本白皮书用其表示AI对学习产出的平均影响。

过程性评价 (Formative Assessment) : 贯穿学习过程的持续性评价, 区别于只看最终结果的标准化考试, 是AI时代评价重构的方向。

AI素养 (AI Literacy) : 理解AI能力边界、能负责任地与之协作的能力, 正成为学生、教师、管理者的必修课。

数字鸿沟 (Digital Divide) : 不同群体在算力、数据、优质模型获取上的差距; 教育AI若不加干预, 可能放大而非缩小这一鸿沟。