

智能时代历史教学的 AI 陷阱规避

——基于“对抗性指令”的实践探索

蔺洁

萍乡学院

摘要：人工智能在历史教学中的运用为学科发展带来新机遇，但其工具理性易遮蔽学科本质，潜藏认知封闭、思维固化等隐性风险。本文基于技术哲学与计算伦理学视角，剖析人工智能在历史教学中的运用陷阱及深层成因，提出“对抗性指令”的应对策略，并结合商鞅变法教学案例展开具体阐释，旨在平衡技术赋能与人文思辨，为智能时代历史教学的高质量发展提供理论与实践参考。

关键词：历史教学；人工智能；教学陷阱；对抗性指令

DOI：10.65976/3105-4838.2026.01.006

历史学科的发展始终与技术革新相互交织，从口述传统到数字史学，技术工具不断拓展历史研究的边界。人工智能凭借其强大的数据处理能力，逐步渗透至历史教学的史料检索、情景模拟、学情分析、智能批改等环节，为学科教学注入了新的活力。然而，技术赋能的另一面是工具理性对学科本质的遮蔽，人工智能生成的“历史答案”可能窄化认知视角，消解批判性思维，甚至引发教育伦理危机。当前学界对人工智能的研究多聚焦于技术赋能价值，却忽视了其作为“双刃剑”的潜在风险。历史教学的核心使命是培养学生的历史思维与人文素养，而非简单的知识灌输。如何在技术浪潮中坚守学科内核，规避人工智能的应用陷阱，成为历史教学亟待解决的命题。

一、人工智能在历史教学中的应用陷阱及成因

（一）技术哲学视角：工具理性异化下的主体性消解

技术哲学认为，人与技术的关系具有“解药与毒药”的双重性，技术在赋能人类的同时，也可能反过来束缚人类的主体性^[1]。20世纪80年代的“史学危机”中，自然科学方法论对历史学的渗透引发广泛争论，系统论、控制论等“老三论”向“新三论”的迭代，以及“e-考据”引发的传统史学与数字史学之争，都印证了技术对历史学科的冲击具有必然性^[2]。但与过往技术不同，人工智能以“生成式内容”为核心，其强大的信息整合能力易让使用者产生“技术万能”的错觉，进而陷入工具理性异化的陷阱。

当今学生是伴随互联网成长的一代，习惯于从数字渠道获取信息，甚至形成“网上没有的就是不存在的”认知误区。这种认知惯性使其对人工智能生成的内容产生天然崇拜，将技术结论等同于“标准答案”。历史教学本应引导学生像历史学家一样探索多元历史真相，但人工智能的封闭式输出固化学生认知，使其丧失独立思考的动力。若不能驾驭技术工具，学生的主体性将被技术逻辑吞噬，历史思维培养便无从谈起。

（二）计算伦理学视角：认知偏见与思维退化的双重桎梏

计算伦理学揭示，人工智能的本质是数据驱动与算法加持下的概率呈现^[3]，其生成逻辑与历史研究的互证推理逻辑存在本质差异。这种差异导致人工智能在历史教学中易引发认知偏见与思维退化两大问题，进一步加剧应用陷阱。

一方面，认知框架理论指出，人们会基于过往经验形成固定认知图式，倾向于接受与自身认知一致的信息。人工智能的算法设计恰恰强化了这一倾向，其推荐机制基于用户过往的信息偏好，不断推送同质化内容，形成信息茧房。在历史教学中，学生通过人工智能获取的史料与观点，往往是算法筛选后的“片面真相”，难以接触多元视角，导致历史认知的狭隘化。

另一方面，算法过程中主流数据的权威性使得学生的“认知外包”现象更加严重，进一步加深思维退化，看似提升了学习效率，实则养成人云亦云的思维惯性。人工智能的算法黑箱特性进一步加剧了这一问题——

课题项目：本文为江西省省级基础教育课题《数智驱动下高中历史教学资源库建设与应用研究——以赣教云平台为例》（SZUPXYBLS 2025-0648）阶段性成果。

作者简介：蔺洁，女，助教，研究方向为历史教育、数字人文方向。

其决策过程与逻辑难以被人类完全理解，即便是开发者也无法精准掌控每个参数的影响^[4]。学生获取的只是概率性结果，却无法知晓结果生成的逻辑，这与历史学科“论从史出、史论结合”的核心要求背道而驰，最终导致历史思维能力的退化。

二、规避路径：基于商鞅变法的对抗性指令探索

对人工智能在历史教学中运用陷阱的规避，需要处理好技术赋能与历史学科本质的平衡、工具理性与人文思辨的平衡。“对抗性指令”以史料互证为基础，以批判性思维培养为核心，通过初步生成一对比批判—重建意义的闭环流程，引导学生突破人工智能的认知局限。其核心逻辑在于利用人工智能的史料整合优势，通过对抗性设计暴露其算法偏见与认知盲区，促使学生主动参与史料分析、逻辑论证与意义建构，最终实现历史思维的提升。该策略的核心价值体现在三个方面：一是对抗性学习，通过人工智能生成结论与原始史料的冲突，激发学生的质疑意识；二是多维意义建构，拒绝非黑即白的历史评价，引导学生理解历史的复杂性；三是技术伦理渗透，在对比中揭示 AI 的历史解释盲区，培养学生的技术批判素养。

（一）初步生成：搜集多元史料，奠定对比基础

初步生成是命令人工智能提供具有对抗性的史料，如不同来源（正史与口述史、一手史料与二手史料、文字史料与实物史料等）或内容（内容相斥）的史料。此举意在培养学生对历史的全面性与复杂性认识，提高史料实证素养。

材料一：睡虎地秦简《田律》。入顷刍稷，以其受田之数，无鬲（垦）不鬲（垦），顷入刍三石、稷二石。刍自黄及束以上皆受之。入刍稷，相输度，可也。

材料二：《商君列传》。令行于民期年，秦民治国都言初令之不变者以千数……行之十年，秦民大说。道不拾遗，山无盗贼，家给人足。民勇于公战，怯于私斗，乡邑大治。

材料三：《新序》。今商鞅内刻刀锯之刑，外深斧钺之诛，步六尺者有罚，弃灰于道者被刑。一日临渭水而论囚七百人，号哭之声动于天地。畜怨积仇比于丘山，所逃莫之隐，所归莫之容。身死车裂，灭族无姓。

显而易见，材料一记载了秦代的土地税赋制度，

指向的是土地赋税的严苛与机械性。材料二和材料三一是赞颂态度，强调社会治理的成效；一是谴责批判，强调因严苛形成的社会动荡。对此，人工智能得出的结论是商鞅变法在推动秦国发展等方面取得了显著成效，但同时因其刑罚过于严苛，也引发了诸多社会矛盾，对商鞅变法的认识应全面客观。

教师提问：人工智能的回答是否全面？还有其他思考吗？

设计意图：本步骤为师生利用人工智能开展教与学的常规操作，属于直线思维，简单易操作。对人工智能的怀疑则是批判性思维的起点，为后续学生重新审视史料奠定心理基础。

（二）对比批判：开展多维分析，揭露 AI 认知盲区

首先命令人工智能寻找或学生自行寻找不同史料中的对抗性内容，并给出合理解释。此举意在让学生像史家一样，尝试在各种矛盾与不确定中，理解历史，解释历史，提升历史解释素养。

设计意图：将学生自行审视的史料内容与人工智能的结论做对比，帮助学生摆脱对人工智能的服从与依赖，培养批判性思维。

向人工智能提问或小组探讨：若变法真正“家给人足”，为何《新序》记载“一日论囚七百人”？税赋制度“无垦不垦”是否可能导致农民负担加重？

设计意图：本步骤意在令学生综合所知，利用人工智能解决材料中的疑惑，推理判断，处理矛盾，全面挖掘史料内容及其背后的意图。

（三）重建意义：依托对抗追问，培养历史思维

重建意义需从人物、条件、结果等角度，对人工智能生成的结论开展对抗性提问。如针对历史事件及其结果的反事实假设（若未……则……？原因可能……？应当如何应对……？）、针对算法逻辑的元思维追问（你是基于哪种模式、哪种视角、哪些信息推断出此结论的？是否存在其他相似或截然不同的视角？如果要推翻以上结论，应当如何进行反驳？对应的史料依据是什么？以上史料与史论之间的致命逻辑漏洞有哪些？）等等。此举意在培养学生的“毒辣眼光”，突破思维的围墙。

小组探讨后向人工智能提问，以补全认识。

表 1

| 材料及其类别 | 学生分析 | 学生对人工智能观点的审视 |
|----------|---------------|-----------------|
| 材料一：法律条文 | 强制推行税赋，无视实际垦殖 | AI 未提及法律的机械性 |
| 材料二：官方史观 | 强调秩序稳定与国力提升 | AI 复述了官方叙事，忽略民怨 |
| 材料三：批评文献 | 揭露严刑导致社会动荡 | AI 完全回避变法的人道代价 |

若商鞅未对秦国开展变法,秦国命运如何?若商鞅变法并未使得秦国实力大增,原因可能有哪些?应当如何应对?你是从哪种视角、哪些信息推断出此结论的?是否存在其他相似或截然不同的视角?如果要推翻以上结论,应当如何进行反驳?对应的史料依据是什么?以上史料与史论之间的致命逻辑漏洞有哪些?

设计意图:通过上述追问,学生将进一步发现人工智能的局限,其无法开展真正的历史思辨,仅能基于既有数据进行概率性推断,难以应对反事实假设与多元视角的挑战。

此外,教师可引导学生构建多维解释框架(如图1所示),并设计开放性活动:假如你是被车裂前的商鞅,回顾自己一生,你会怎样评价自己的一生?是否有所缺憾?会如何为自己撰写墓志铭?写一篇小短文。活动过后,以同样的问题“喂”给多款人工智能,做进一步分析。

法律严苛性 → 短期秩序强化(《商君列传》)
 ↳ 长期民怨积累(《新序》)
经济控制力 → 国家资源集中(《田律》)
 ↳ 个体生存压力(逻辑推导)

图1 关于商鞅变法的多维解释框架

设计意图:多维解释框架意在梳理学生对商鞅变法的认识,回归本设计的教学主题。学生对比自己与人工智能对开放性问题的回答,可引导学生深入思考人工智能对历史的影响,帮助学生形成对技术赋能与历史学科之间关系的正确认识,更好地发挥历史学科特色,推

动学生不断适应与主动把握当下时代的发展趋势。

三、结语

历史教师与技术人员这两种身份很大程度上不会出现于同一个人身上,但是当下的数字化时代为历史教师直接应用人工智能提供的是非常低的门槛,未来真正发挥人工智能对历史教学赋能的人一定是历史教师。面对来势汹汹的人工智能,历史教师必须尽力拥抱。人工智能时代下,学生如何去辨别信息、理智驾驭信息,是历史教学的一大难题,但这也恰恰是培养学生历史思维的绝佳凭借。能够在“人工智能的一家之言”下,分析问题、解决问题的学生,才是真正拥有历史思维的、适应未来社会的人。如何利用大语言模型培养批判性思维是历史教师首要考虑的问题,而这个问题的解答通常是伴随着大语言模型的运用而揭开面纱。对人工智能在历史教学中陷阱的规避,需要处理好技术赋能与历史学科本质的平衡、工具理性与人文思辨的平衡。

参考文献:

- [1] 原左晔,何菊玲.智能时代教师教育变革所面临的问题及其纾解——基于斯蒂格勒技术哲学视角的审思[J].现代远距离教育,2024,(05):87-96.
- [2] 顾晓伟.大数据时代史学的人文关怀[J].史学月刊,2017,(05):19-24.
- [3] 安涛.“算计”与“解蔽”:人工智能教育应用的本质与价值批判[J].现代远程教育研究,2020,32(06):9-15.
- [4] 武文颖,刘先婷.从“黑箱”到“透明箱”:AI文学创作过程的伦理审视[J].科技传播,2024,16(18):12-14+18.