Advances in Education 教育进展, 2026, 16(9), 244-255
Published Online September 2026 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2026.1691899

Hans汉斯

离散数学教学中Chain-of-Thought可验证推理能力的现状与改革方向

赵　帅*，汪云路，许艳萍，吕秋云

杭州电子科技大学网络空间安全学院，浙江　杭州

收稿日期：2026年7月31日；录用日期：2026年8月29日；发布日期：2026年9月4日

摘　要

随着DeepSeek、GPT等具备Chain-of-Thought (CoT)推理能力的生成式AI的普及，离散数学课程教学面临一个前所未有的教学挑战：学生应对课程作业能够轻易获取AI生成的“分步推理”的答案，但其独立判断这些推理是否正确的核心能力是否得到培养，已成为亟待关注的教学问题。本研究以2026春季学期29名离散数学学生为调查对象，通过问卷调查分析方法，系统考察了学生在AI辅助学习中的使用情况、推理能力自评及对CoT可验证推理的认知状况。研究发现：(1) 96.6%的学生在离散数学课程学习过程中使用过AI工具辅助学习，其中79.3%达到经常或偶尔使用频率，AI已深度嵌入学生的学习过程；(2) 学生在“发现证明漏洞”和“理解分步构造”两个维度自评较高(M = 4.00、4.07)，但在“判断AI生成证明是否正确”(M = 3.76)和“识别AI回答中的错误”(M = 3.86)两个维度的自评显著偏低；(3) 高频使用AI的学生在所有推理能力维度上的自评均高于低频使用者，尤其在“独立完成证明能力”上的差值达0.45，但学生对AI输出的批判性判断能力并未随使用频率增加而同步提升。基于调查结果与认知负荷理论、元认知理论的分析框架，本文提出离散数学教学亟需从“答案导向”转向“CoT可验证推理导向”，并构建“独立构造CoT推理链、AI辅助验证、人机协同修正、批判性判断”的四阶段教学框架，以重建学生在AI时代的数学推理主体性，为离散数学课程应对AI时代的教学挑战提供方向性参考。

关键词

离散数学，Chain-of-Thought，可验证推理

Current State and Reform Directions of Chain-of-Thought Verifiable Reasoning Ability in Discrete Mathematics Teaching

Shuai Zhao*, Yunlu Wang, Yanping Xu, Qiuyun Lyu

School of Cyberspace, Hangzhou Dianzi University, Hangzhou Zhejiang

*通讯作者。

文章引用: 赵帅, 汪云路, 许艳萍, 吕秋云. 离散数学教学中 Chain-of-Thought 可验证推理能力的现状与改革方向[J]. 教育进展, 2026, 16(9): 244-255. DOI: 10.12677/ae.2026.1691899

Received: July 31, 2026; accepted: August 29, 2026; published: September 4, 2026

Abstract

With the widespread adoption of generative AI models such as DeepSeek and GPT that possess Chain-of-Thought (CoT) reasoning capabilities, discrete mathematics teaching now faces an unprecedented challenge: while students can easily obtain AI-generated "step-by-step reasoning" answers for course assignments, whether their core ability to independently judge the correctness of such reasoning is being cultivated has become a pressing instructional concern. This paper investigates 29 discrete mathematics students from the 2026 spring semester as survey subjects. Through questionnaire-based analysis, it systematically examines students' usage patterns of AI-assisted learning, self-evaluated reasoning abilities, and awareness of CoT verifiable reasoning. The findings reveal that: (1) 96.6% of students have used AI tools to assist their learning in discrete mathematics, with 79.3% using them frequently or occasionally. Thus, AI has become deeply embedded in students' learning processes; (2) students rated themselves relatively high in "identifying proof gaps" (M = 4.00) and "understanding step-by-step construction" (M = 4.07), but rated themselves significantly lower in "judging whether AI-generated proofs are correct" (M = 3.76) and "identifying errors in AI responses" (M = 3.86); (3) high-frequency AI users scored higher on self-evaluations across all reasoning dimensions compared to low-frequency users, with a notable difference of 0.45 on "independent proof construction ability," yet students' critical judgment of AI-generated output did not improve correspondingly with increased usage frequency. Drawing on the analytical frameworks of Cognitive Load Theory and Metacognitive Theory, this paper argues that discrete mathematics education urgently needs to shift from an "answer-oriented" approach to a "CoT verifiable reasoning-oriented" paradigm. To reestablish students' mathematical reasoning agency in the AI era, it proposes a four-stage instructional framework, *i.e.* independently constructing CoT reasoning chains, AI-assisted verification, human-AI collaborative revision, and critical judgment, thereby providing directional guidance for discrete mathematics courses to address pedagogical challenges in an AI-driven academic environment.

Keywords

Discrete Mathematics, Chain-of-Thought, Verifiable Reason

1. 引言

离散数学是计算机科学的核心基础课程，其本质任务是培养学生的数学推理能力。从命题逻辑到谓词逻辑、从数学归纳法到图论证明、从组合计数到代数结构，学生需要在这一课程中建立“什么是严格证明”的认知框架。然而，传统离散数学教学长期面临一个深层矛盾：学生能够模仿教师讲授的证明方法完成习题，但面对全新的证明问题时，往往既无法独立构造证明，也无法有效判断一个证明的正确性。离散数学课程存在“概念密集、理论抽象、知识跨度大”等特点[1]，传统的课程教学模式难以满足对于学生数学思维的培养要求。

进入大模型时代后，这一矛盾呈现出新的形态。2024 年，Google DeepMind 正式发布的 AlphaProof 系统将形式化证明语言 Lean 与强化学习相结合，成功解决了多道国际数学奥林匹克金牌级别的题目，展

示了 AI 进行严格数学推理的可行性[2] [3]。同年，OpenAI 发布 o1/o3 系列推理模型，引入 Chain-of-Thought (CoT)推理机制。在输出答案之前，模型会经历一个较长的内部思维链过程，将复杂问题分解为若干可追溯的中间步骤。国内方面，DeepSeek 团队发布的 R1 系列推理模型在数学推理任务上展现出与 OpenAI o1 相当的能力，并完全开源其推理思考过程。这些进展的共同特征在于：推理智能体能够生成可追溯、可验证的链式思维推理链条[4]。

Chain-of-Thought 推理的核心机制在于将复杂问题分解为一系列中间推理步骤，每一步都建立在前一步的基础之上。这一机制与离散数学证明教学的本质要求高度契合，证明本身就是将一个复杂的命题分解为若干可验证的逻辑步骤。然而，技术突破的另一面是教育挑战。当学生可以轻松让 DeepSeek 或 ChatGPT 生成一个看似合理的“分步证明”时，他们是否还具备独立构造证明的能力？更关键的是，他们是否还具备判断 AI 生成证明是否正确的能力？张妙雨等人研究指出“传统考核中高分的学生常因缺乏对结构的直觉而在 AI 幻觉面前失效”[5]。这一现象在离散数学的证明教学中尤为突出，学生看到了 AI 展示的“推理步骤”，却难以独立判断这些步骤之间的逻辑连接是否成立。

在教育应用层面，国际高校已经开始了一些富有启发性的探索[6]。斯坦福大学的 CS109 课程引入了“AI 辅助证明检查”环节，引导学生先独立完成证明，再借助 AI 工具进行验证和修正[7]。有研究发现，“学生在批评不完美的 AI 生成论证时，往往比批评同学写的草稿更加投入，这使得这些工具在培养批判性思维方面具有独特的有效性”[8]。国内高校近年来也在积极探索“人工智能+”教学改革，如赵鑫等人研究提出了基于知识图谱与多专家智能体的创新教学模式[9]，李琳等人尝试将程序设计的实践教学环节引入离散数学课程[10]。然而，现有实践多集中于课程内容生成、习题自动批改、智能答疑等辅助性环节，在培养学生“CoT 可验证推理”能力这一核心维度上，仍缺乏系统性的教学研究和实证数据。

CoT 的核心特征是可追溯、可分解、可验证，恰恰对应了离散数学证明教学的本质要求。然而，当前国内在培养学生“CoT 可验证推理”能力方面尚缺乏系统性的教学研究和实证数据。多数“AI + 数学教学”的研究仍集中于 AI 辅助课程内容生成、习题自动批改、智能答疑等辅助性环节，而忽略了最核心的问题[5]。当 AI 能够模拟甚至超越人类的推理过程时，学生自己的推理能力该如何定位和培养？

本研究以某高校 2026 春季学期 29 名离散数学学生为调查对象，通过 29 份有效问卷，系统考察了学生在 AI 辅助学习中的真实行为模式、推理能力自评状况以及对 CoT 可验证推理的认知水平，旨在回答以下三个问题：(1) AI 工具在离散数学学习中的渗透程度和使用特征是什么？(2) 学生在“使用 AI 辅助证明”与“独立判断证明正确性”之间是否存在能力落差？(3) 该落差对离散数学教学改革提出了怎样的启示？

2. 调查设计

2.1. 调查对象

调查对象为某高校网络空间安全学院信息安全专业 2026 春季学期一个自然班的 29 名学生，均修读《离散数学》课程。该课程涵盖命题逻辑、谓词逻辑、集合与关系、组合计数、图论五个核心模块，以证明推理能力培养为核心教学目标。班级总人数 34 人，回收问卷 29 份，有效回收率 85.3%。

2.2. 调查工具

调查采用自编问卷，如表 1 所示，包含四个部分：(1) AI 工具使用频率与场景(单选题 + 多选题)；(2) 推理能力自评量表(9 道五级李克特量表题[11]，从“非常不同意”到“非常同意”分别赋值 1~5 分)；(3) 使用方式偏好(单选题)；(4) 开放式问题(4 道简答题)。问卷通过《学习通》在线平台发放，学生匿名填写。

Table 1. Structure of the survey questionnaire
表 1. 问卷调查结构

问卷部分	题型	内容描述	对应题项
AI 工具使用情况	单选 + 多选	AI 工具使用频率、使用场景、提问类型	Q1~Q3
推理能力自评量表	五级李克特量表(1 = 非常不同意，5 = 非常同意)	9 道题，覆盖 CoT 可验证推理的接受与理解、独立推理能力与判断力、批判性 AI 使用素养三个维度	Q4~Q12
使用方式偏好	单选	对 AI 辅助方式的倾向性选择	Q13
开放式问题	简答	最大收获、遇到问题、调整方向、其他建议	Q14~Q17

为系统测量学生在“使用 AI”与“判断 AI”之间的能力落差及其结构特征，如表 2 所示，量表中的 9 道五级李克特量表题目在设计上对应三个理论维度：CoT 可验证推理的接受与理解(Q4 发现漏洞、Q5 理解分步构造、Q6 清晰表达推理思路)、独立推理能力与判断力(Q7 独立完成证明、Q8 判断 AI 证明正确性、Q9 证明信心提升)以及批判性 AI 使用素养(Q10 引入 AI 有益、Q11 识别 AI 错误、Q12 先独立再验证有帮助)。这一设计旨在从多个角度捕捉学生在“使用 AI”与“判断 AI”之间的能力结构，三个维度之间形成了从“接受理解”到“独立应用”再到“批判反思”的递进关系。

Table 2. Three theoretical dimensions and corresponding items of the self-assessment scale of reasoning ability
表 2. 推理能力自评量表的三个理论维度与对应题项

理论维度	核心内涵	对应题项
CoT 可验证推理的接受与理解	衡量学生对 AI 辅助的感知有效性，能否借助 AI 发现推理漏洞、理解分步构造、清晰表达思路。反映学生“消费”CoT 推理链的能力。	Q4~Q6
独立推理能力与判断力	衡量学生脱离 AI 辅助后的独立推理水平，以及对 AI 输出正确性的判断能力。反映学生能否从“消费 CoT”走向“验证 CoT”。	Q7~Q9
批判性 AI 使用素养	衡量学生对 AI 辅助的反思性态度，能否在认同 AI 价值的同时识别其错误，认同“先独立再验证”的教学安排。反映学生批判性使用 AI 的意识。	Q10~Q12

2.3. 数据分析方法

本研究采用定量与定性相结合的混合分析方法，对回收的 29 份有效问卷进行系统分析。

在定量分析方面，首先对原始数据进行编码处理：将单选题和量表题按 A = 1、B = 2、C = 3、D = 4、E = 5 赋值，多选题按各选项是否被选择进行 0/1 二值编码，开放式问题单独整理为文本语料库。在此基础上开展以下三类统计分析。

(1) 描述性统计：计算各单选题和多选题的频次与百分比，以描绘 AI 工具的使用特征；计算各量表题的均值与标准差，以评估学生在各维度上的总体自评水平。

(2) 组间均值比较：将学生按 AI 使用频率分为高频组(“经常使用”+“偶尔使用”，23 人)与低频组(“很少使用”+“从未使用”，6 人)，计算两组在各量表题上的均值差值，以考察使用频率与推理能力自评之间的关联。

(3) 使用偏好分析：对使用方式偏好的选项分布进行频次统计，以了解学生对不同 AI 使用策略的选择倾向。

在定性分析方面，对开放式回答采用主题编码分析法。识别文本中反复出现的主题与关键词，形成初始编码方案；随后合并相近类别、调整不一致的编码，直至形成最终编码框架。编码完成后，统计各

主题的出现频次，并选取最具代表性的学生原话作为典型例证。

最后，将定量数据与定性数据进行交叉验证：将量表题各维度的得分情况与开放式问题中学生反映的收获与问题主题进行对照分析，判断两类数据是否指向一致的结论，从而实现量化数据与质性数据的相互补充与印证。

3. 调查结果

3.1. AI 工具的使用现状

根据问卷调查分析结果，AI 工具已深度嵌入离散数学的学习过程。96.6% (28/29)的学生在本学期使用过 AI 工具辅助学习，其中“经常使用(每周 3 次以上)”占 37.9% (11 人)，“偶尔使用(每周 1~2 次)”占 41.4% (12 人)，“很少使用(整个学期少于 5 次)”占 17.2% (5 人)，“从未使用过”仅占 3.4% (1 人)。经常与偶尔使用者合计占比 79.3%，说明 AI 辅助学习已成为当前该班级学生的普遍行为而非个别现象。

在使用场景方面(多选)，占比从高到低依次为：复习备考(72.4%，21 人)、验证自己证明思路的正确性(69.0%，20 人)、课堂学习遇到疑问时(65.5%，19 人)、完成证明题作业时(58.6%，17 人)。值得注意和欣慰的是，使用场景为“验证自己证明思路的正确性”的比例位列第二，表明学生并非单纯将 AI 作为“答案获取工具”，而是具有一定程度的“验证”意识，他们试图用 AI 来检查自己的推理是否正确。然而，这种“验证”本质上是一种对 AI 输出的依赖式验证，而非基于自身推理能力的独立验证，学生将验证的任务外包给了 AI，而非内化为自己的思维习惯。

在提问类型方面(多选)，“请帮我检查这个证明是否正确”(62.1%，18 人)和“请一步一步推理这道证明题”(58.6%，17 人)是最常见的两类提问，合计超过 60%。这说明大部分学生已经在直觉层面理解到了 CoT 推理的核心，他们希望看到“一步一步”的推理过程，也希望 AI 来“检查”自己的推理链。然而，这种接触是浅层的、被动的，学生是在“消费”AI 生成的 CoT 推理链，而非“建构”自己的 CoT 推理链。AI 的 CoT 推理本质上是“自回归”生成的，“顺着自己的思路往下说，更在乎表达流畅，反而容易忽略逻辑严谨性”[12]-[14]。学生如果只消费而不审视 AI 的 CoT 输出，AI 所呈现的流畅表达便可能成为一种认知陷阱，使学生误将“表达连贯”等同于“逻辑正确”。

3.2. 推理能力自评量表分析

如表 3 所示，9 道量表题的描述性统计结果从均值与标准差两个层面，呈现了学生在三个理论维度(CoT 可验证推理的接受与理解、独立推理能力与判断力、批判性 AI 使用素养)上的自评水平。基于 9 道量表题的得分分布，本文从三个角度对学生的自评结果进行分析：学生对 AI 辅助的接受与感知效果、对 AI 输出的独立判断能力，以及脱离 AI 后的独立证明水平，具体展开如下：

Table 3. Data of self-assessment results on reasoning ability
表 3. 推理能力自评结果数据

题项	题目简述	均值 (Mean, M)	标准差 (Standard Deviation, SD)
Q4	发现证明思路中的漏洞或错误	4	0.54
Q5	理解“如何分步构造证明”	4.07	0.53
Q6	更清晰地表达推理思路	4.17	0.47
Q7	无 AI 辅助时也能独立完成证明	3.9	0.62

续表

Q8	能够判断 AI 生成证明是否正确	3.76	0.69
Q9	相比学期初，证明能力更有信心	4.07	0.53
Q10	在证明学习中引入 AI 有益	4.21	0.49
Q11	识别 AI 回答中的错误或不严谨	3.86	0.58
Q12	“先独立构造再验证”的做法有帮助	3.93	0.53

第一，学生在“接受与理解 AI 辅助”维度的评价普遍较高。Q10 (引入 AI 有益，M = 4.21)和 Q6 (更清晰表达推理思路，M = 4.17)得分最高，Q5 (理解分步构造，M = 4.07)和 Q9 (证明信心提升，M = 4.07)紧随其后。这表明学生普遍认可 AI 辅助的价值，并感知到自己在“分步推理”和“思路表达”两个核心能力上有所提升。值得注意的是，CoT 推理机制的两个核心特征，“分步构造”和“思路表达”，恰好是学生感知提升最明显的地方，这说明 AI 的 CoT 输出确实在一定程度上起到了示范作用。

第二，学生在“独立判断 AI 输出”维度的评价显著偏低。Q8 (判断 AI 生成证明是否正确，M = 3.76)和 Q11 (识别 AI 回答中的错误，M = 3.86)是所有量表中得分最低的两项。这一差值(Q10 的 4.21 与 Q8 的 3.76 相差 0.45)揭示了一个结构性矛盾：学生越是依赖 AI 来“验证”自己的推理，就越缺乏独立判断 AI 输出是否正确的能力。他们能够“消费”CoT 推理链，却无法“验证”CoT 推理链，这正是 CoT 可验证推理能力缺失的直接体现。从认知负荷理论的视角来看[15] [16]，学生将“验证”这一元认知任务外包给 AI 后，其自身的元认知监控能力，即“监控自己的推理过程并判断其是否正确”的能力，未能得到充分发展。学生能够理解 AI 展示的每一步“是什么”，却无法独立判断每一步“为什么成立”以及“是否还有遗漏”。

第三，学生在“独立完成证明”维度的得分处于中等偏下水平。Q7 (无 AI 辅助独立完成证明，M = 3.90)在所有量表中位列倒数第三，仅高于 Q8 和 Q11。这意味着，尽管 AI 辅助提升了学生的分步推理意识和表达清晰度，但这种提升尚未充分转化为“脱离 AI 后的独立证明能力”，学生在 AI 的“脚手架”下能够走得更远，但撤去脚手架后，他们的独立行走能力依然有限。这一发现与“脚手架理论”的核心洞见一致：教学支架的价值在于最终撤除，如果学生始终依赖支架而未能将支架所支撑的能力内化，则教学未能实现其根本目标[17] [18]。

3.3. 使用频率与推理能力的关联分析

将学生按 AI 使用频率分为高频组(“经常使用” + “偶尔使用”，23 人)与低频组(“很少使用” + “从未使用”，6 人)，表 4 比较两组在 9 项能力自评上的差异：

高频组在所有 9 个维度上的自评均高于低频组，其中差值最大的是 “独立完成证明能力”(Q7, +0.45)和“发现证明漏洞”(Q4, +0.38)。这一趋势提示：更频繁地使用 AI 辅助学习的学生，在自我感知的推理能力和独立证明信心上具有一定优势，这与“AI 作为推理伙伴”的定位一致。然而，在“判断 AI 证明正确性”(Q8)和“识别 AI 错误”(Q11)两个维度上，高频组的优势并不突出(差值分别为 0.27 和 0.19)，说明即使高频使用 AI，学生的批判性判断能力也未必随之同步发展——这可能源于当前学生使用 AI 的方式多停留在“接收反馈”而非“审视输出”层面。这一发现与元认知理论的核心观点相呼应：元认知能力(即“对自身认知过程的认知和监控”)不会随着外部工具的使用频率增加而自动发展，它需要专门的教学设计和刻意训练。

Table 4. Data of correlations between on reasoning ability
表 4. 使用频率与推理能力的关联分析结果数据

题项	高频组均值	低频组均值	差异
Q4 (发现证明漏洞)	4.05	3.67	+0.38
Q5 (理解分步构造)	4.09	3.83	+0.26
Q6 (清晰表达推理思路)	4.18	4.00	+0.18
Q7 (独立完成证明能力)	3.95	3.50	+0.45
Q8 (判断 AI 证明正确性)	3.77	3.50	+0.27
Q9 (推理信心提升)	4.09	3.83	+0.26
Q10 (引入 AI 有益)	4.23	4.17	+0.06
Q11 (识别 AI 错误)	3.86	3.67	+0.19
Q12 (先独立再验证有帮助)	3.95	3.83	+0.12

3.4. 使用方式偏好

除了自评量表中学生对 AI 辅助效果的感知评价外，本文在题项 Q13“你更倾向于以下哪种使用方式”一题中，考察了学生对 AI 工具的态度，折射出其元认知判断力。在该题中，选择“先自己思考并写出完整证明，再用 DeepSeek 验证和修正”的占 24.1% (7 人)，“遇到不会的题直接问 DeepSeek 要答案”的占 6.9% (2 人)，“两者结合，视题目难度而定”的占 69.0% (20 人)。

近七成学生选择了“视题目难度而定”的灵活策略，选择“直接要答案”的仅 2 人。这一结果可从两个角度加以解读。积极的一面是，学生并非简单地将 AI 作为“抄答案”工具，而是具备一定的元认知判断力，能够根据题目难度调整自己的求助策略，这反映了对自身能力边界的初步识别。审慎的一面是，“视难度而定”也可能成为回避深度思考的策略，当学生以“这道题太难了”为由选择求助 AI 时，他们实际上放弃了在“最近发展区”内通过适度挑战实现能力成长的机会[19]。两种解读并非互斥，在实际学习情境中可能同时存在。对学生究竟是“策略性求助”还是“策略性回避”，需要结合其具体使用行为(如是否先尝试独立构造推理链)加以判断，仅凭偏好选项难以做出确定性论断。

3.5. 开放式问题

为了进一步探究学生在“CoT 可验证推理的接受与理解”、“独立推理能力与判断力”和“批判性 AI 使用素养”三个理论维度上自评量表刻画结果背后的深层次原因，以及为后续教学改革提供现象学依据，本章在自评量表基础上加入了开放式问题，表 5 展示了问题设置的对应目标，基于主题分析编码，将学生答案中的文本数据转化为可量化的主题词，具体分析如下。

Table 5. Objectives for open-ended questions
表 5. 开放问题目标设置

题项	问题内容	问题目标
Q14	使用 DeepSeek 等 AI 辅助工具学习离散数学，你最大的收获是什么？	收集学生对 AI 辅助价值的感知与评价，为量表发现提供补充说明。
Q15	使用 AI 辅助工具过程中遇到了哪些问题或不便之处？	揭示学生使用 AI 过程中的障碍与风险，弥补量表在解释力上的局限。

续表

Q16	如果有机会重新学习这门课，你会在哪些方面调整使用 AI 的方式？	捕捉学生的元认知反思与自主学习意识，考察学生对自身使用策略的评价与调整意愿。
Q17	你想对老师说的其他建议或想分享的体验。	收集学生自发提出的教学改进建议，补充前序问题未覆盖的经验信息。

Q14“最大的收获”(29 人作答，有效回答 25 条)：

编码显示，收获集中在三个主题：理清推理/证明思路(8 人次，如“AI 能把抽象的逻辑推导拆成步骤，帮我理解‘为什么想到这一步’”)、高效理解概念/定理(7 人次，如“解释不会的概念”“理解晦涩定理”)和提升学习效率(5 人次)。值得注意的是，没有一条回答提及“我学会了判断 AI 给的证明对不对”，学生的收获感知集中在“AI 帮我理解”，而非“我能判断 AI”。这一发现揭示了学生对 AI 辅助的认知存在一个盲区：他们将 AI 视为知识的传递者而非思维的对话者。

Q15“遇到的问题”(29 人作答，有效回答 22 条)：

最主要的问题是“AI 给出错误/假答案”(8 人次)，其次是“AI 理解/识别能力差”(6 人次，如“识图不便”“公式 OCR 识别不准”)和“过度依赖、自主思考减少”(4 人次)。有学生写道：“AI 容易顺着我错误的思路走，不及时指出逻辑漏洞”，这正是 CoT 可验证推理缺失的危险信号：AI 不仅没有纠正学生的错误推理，反而在顺着错误的方向“协同推理”，使学生在错误路径上越走越远。大模型的 CoT 推理本质上是“自回归”生成，“顺着自己的思路往下说，更在乎表达流畅，反而容易忽略逻辑严谨性”。当这种“自回归”特性遇到学生已有的错误推理倾向时，就形成了“错误放大”的恶性循环。

Q16“调整使用方式”(29 人作答，有效回答 18 条)：

学生最希望调整的方向是“先独立思考再使用 AI”(5 人次)，其次是“更多用于理解概念/思路而非直接做题”(4 人次)。有学生明确提出：“把 AI 的解释和教材对照，标记差异，培养信息甄别能力”，这一表述精准地指向了 CoT 可验证推理的核心：不仅要知道 AI“怎么想”，还要有能力判断 AI“想得对不对”。还有学生提出“专门用 AI 找反例，训练批判性思维”，这体现了部分学生对 AI 工具的深度思考和主动规划意识。

4. 讨论

4.1. “CoT 消费者”与“CoT 验证者”之间的能力落差

调查结果揭示了一个关键矛盾：学生在使用 AI 的过程中高频接触 CoT 推理链(62.1%的学生曾让 AI“一步一步推理”证明题)，但他们并未因此获得独立验证 CoT 推理链的能力。学生更像是 CoT 的“消费者”，接收 AI 展示的推理步骤、借助 AI 理解概念、用 AI 验证自己的答案，却尚未成为 CoT 的“验证者”，还不能够独立判断一个 CoT 推理链的每一步是否逻辑严谨、是否完整无漏。这一落差可以从认知负荷理论得到解释。CoT 推理链将复杂的证明过程分解为若干中间步骤，本质上是一种“认知卸载”[20]，将部分推理负担从学生的工作记忆中转移到外部媒介(AI 生成的文本)上。这种卸载在短期内提升了学生的理解效率(正如 Q6 和 Q9 的高分所示)，但长期而言，如果学生始终处于“消费”而非“建构”CoT 的状态，他们的元认知监控能力，即“监控自己的推理过程并判断其是否正确”的能力，将得不到充分发展。学生能够理解 AI 展示的每一步“是什么”，却无法判断每一步“为什么成立”以及“是否还有遗漏”。

从元认知理论的视角来看，推理能力包含两个层面：一是认知层面的推理执行能力(即“会推理”)，二是元认知层面的推理监控能力(即“知道自己推理得对不对”)。AI 的 CoT 输出在认知层面提供了丰富

的示范，学生看到了“如何分步推理”。但是，在元认知层面，AI 的介入反而可能产生负面影响，学生将“验证”这一元认知任务外包给了 AI，自身的监控能力未得到锻炼。这解释了为什么学生在“理解分步构造”上得分较高(M = 4.07)，却在“判断 AI 证明正确性”上得分较低(M = 3.76)，他们学会了“跟着 AI 的步骤走”，却没有学会“判断这些步骤对不对”。

4.2. “AI 顺向推理”陷阱与批判性思维的缺失

开放式问题中一个值得警惕的现象是：多名学生反映“AI 容易顺着我错误的思路走”。这意味着，当学生向 AI 提交一个存在逻辑漏洞的推理链时，AI 非但没有指出漏洞，反而沿着错误的方向继续“推理”，生成一个表面上自洽实则错误的完整证明。这一现象的本质是：AI 的 CoT 推理链是“结果导向”的，它致力于生成一个连贯的推理过程，而非致力于“发现并纠正”推理中的错误。

对于缺乏独立判断能力的学生而言，这种“顺向推理”构成了严重的安全隐患。学生不仅无法识别 AI 输出中的错误(Q8 均值仅 3.76)，还可能因为 AI“分步展示了推理过程”而误以为该证明是严谨的，AI 的 CoT 形式给了学生一种虚假的“可验证性”错觉。学生在面对 AI 生成的数学证明时，往往“因缺乏对结构的直觉而在 AI 幻觉面前失效”。“AI 幻觉”在本研究调查中被学生识别为一个显著问题(8 人次提及“AI 给假答案”)，但其危害不仅在于答案的错误，更在于 CoT 形式所营造的虚假确定性。从技术层面来看，大语言模型的幻觉源于其概率语言模型本质，模型通过最大化序列生成概率来逐一生成词元，而非通过逻辑规则进行推演。当 AI 的“流畅表达”与学生的“判断力缺失”相遇时，就形成了一个危险的教育场景：学生越信任 AI 的 CoT 输出，就越少动用自身的批判性思维去审视它，最终导致判断能力的进一步退化。自回归的 CoT 推理存在“单向归纳偏置”，早期步骤中的一个逻辑错误会在后续生成中被不断放大，最终破坏整个推理链的可靠性。对于缺乏独立判断能力的学生而言，AI 输出的 CoT 形式本身构成了一个“可验证性错觉”，推理链的流畅性遮蔽了其可能存在的逻辑断裂。因此，培养学生在人机交互情境下的批判性信息素养，不仅是教育问题，更是应对生成式 AI 技术特征的必然要求。

4.3. 从“答案导向”到“CoT 可验证推理导向”的教学转向

传统离散数学教学遵循“教师讲授证明方法、学生模仿练习、教师批改反馈”的模式，其隐含假设是：推理能力的培养路径是“从模仿到创造”。在大模型时代，这一路径被 AI 改变，学生不再依赖“模仿”人类的证明，他们可以直接“调用”AI 生成的证明，该发现对离散数学的教学提出了根本性的挑战。

同时，调查结果清晰地表明，AI 的普及并未自动提升学生的推理能力，反而可能侵蚀学生的独立判断力。学生越是依赖 AI 来“验证”自己的推理，就越缺乏独立判断 AI 输出是否正确的能力。这一悖论要求离散数学教学进行一次根本性的范式转换：从“教会学生如何证明”转向“教会学生如何验证证明”，从培养学生成为“证明者”转向培养学生成为“CoT 可验证推理的鉴赏者”。

具体而言，离散数学教学应构建以 CoT 可验证推理为核心的“四阶段”教学框架：

第一阶段(独立构造)：学生独立将证明问题分解为若干可追溯、可验证的中间推理步骤，形成自己的 CoT 推理链。这一阶段旨在培养学生“在没有 AI 辅助的情况下分步思考”的能力，是 CoT 可验证推理能力的基础。教师在此阶段的作用是提供“分步思考”的示范和脚手架，但必须逐步撤除，让学生最终能够独立完成 CoT 推理链的构造。在该阶段，教师重视传统离散数学教学方式，依然需要对 AI 工具的使用做一定的限制。

第二阶段(AI 辅助验证)：学生将自己的 CoT 推理链提交给 AI，请 AI 检查每一步的逻辑正确性。但重点不在于获取 AI 的“正确答案”，而在于对比自己的推理链与 AI 生成的推理链之间的差异。这一阶段的核心教学目标不是“让 AI 告诉我对不对”，而是“让我看到 AI 的推理链与我的推理链有什么不同”，

通过对比激发学生的认知冲突，从而促进深度思考。

第三阶段(人机协同修正)：学生基于对比结果，修正自己的推理链，并在这一过程中建立“判断推理链正确性”的内在标准。教师在此阶段应引导学生追问三个问题：AI 的推理链比我好在哪？我的推理链比 AI 好在哪？如果 AI 的推理链有错误，我凭什么能发现它？这三个追问分别指向“学习 AI 的优点”、“建立自己的推理自信”和“培养批判性判断力”三个教学目标。

第四阶段(批判性判断)：学生最终需要独立判断 AI 的修正建议是否正确？自己的修正是否完备？推理链中是否还存在未被发现的漏洞？这一阶段的终极目标是让学生成为“能够对任何 CoT 推理链(无论来自人类还是 AI)进行严格验证”的独立判断者。

该教学框架的核心逻辑在于：将 AI 从“答案提供者”重新定位为“推理对照者”，AI 的价值不在于替学生完成推理，而在于提供一个可供对照、批判和超越的“推理参照系”。学生通过与 AI 的推理链进行对比，逐渐建立自己对正确推理的判断标准。采用苏格拉底方法、CoT 推理和形成性反馈相结合的干预方式，对大学生的数学学习具有积极效果，这为四阶段教学框架的有效性提供了理论支持[21]。此外，关于证明助手(Proof Assistant)的教育研究也为本研究提供了参照。Hanna 等人在本科数学教学中的研究表明，证明助手的价值不在于替代推理，而在于提供一个“即时验证反馈”的环境，使学生在迭代修正中发展证明直觉[6]。本研究的四阶段框架中，AI 的 CoT 输出作为“对照推理链”，其功能类似于证明助手的验证反馈，引导学生发现自身推理链的不足，而非直接给出正确答案。

更进一步地，本文提出的四阶段教学框架本质上是一种特殊形态的人机交互与协同学习，学习者与 AI 之间形成多轮对话式的推理共建关系。在人机交互(HCI)研究领域，学界长期关注“透明性”与“可解释性”问题，用户需要理解系统为何给出特定输出，才能形成有效的信任判断[22]。本研究中学生“无法判断 AI 生成证明是否正确”(Q8, M = 3.76)的现状，恰恰暴露了当前 HCI 在教育场景中的一个薄弱环节：推理智能体虽然“展示”了推理过程，但其内部的概率生成逻辑与人类可理解的逻辑推演之间仍存在“解释鸿沟”。从计算机支持的协作学习(CSCL)视角来看，CoT 可验证推理教学可被理解为一种“人机协作学习”的新形态。CSCL 的价值不仅在于知识的社会建构，还在于通过对话与互动促进学习者的元认知发展[23]。当 AI 成为学习者的“对话伙伴”而非“答案提供者”时，人机之间形成了一种“扩展的协作学习共同体”。本研究四阶段框架中的“人机协同修正”环节，正是这一理念的操作化实现。

5. 改革方向与教学建议

基于上述调查结果与理论分析，本文提出离散数学教学改革的六个方向性建议：

第一，将“CoT 可验证推理”确立为离散数学教学的核心目标之一。当前离散数学的教学目标多聚焦于“学生能够掌握各模块的证明方法”，建议增加一条明确的目标维度：“学生能够独立判断一个 CoT 推理链的逻辑严谨性与完整性”。这一目标的实现需要贯穿整个课程的教学与评价。具体而言，可在课程大纲中明确列出这一目标，并在每章节的教学中通过具体的证明题目来训练和评估。

第二，设计“人机推理链对比”的课堂活动。选取典型的证明题目，让学生先独立写出自己的 CoT 推理链，再展示 AI 生成的 CoT 推理链，组织学生对比分析两者的异同，谁的推理链更完整？谁的推理链存在跳跃？谁的推理链用了不同的策略？通过这种对比，学生将逐渐建立起对“推理链质量”的判断力。具体操作上，可设计为“15 分钟独立写作 + 5 分钟 AI 生成展示 + 15 分钟小组讨论 + 5 分钟全班分享”的课堂流程，既保证了学生的独立思考空间，又充分利用了 AI 的对比价值。

第三，在作业评价中纳入“推理链验证”维度。传统的作业评价只关注“最终答案是否正确”，建议增加“推理链自我验证报告”。要求学生提交证明的同时，提交一份对自己的 CoT 推理链的“验证报告”，指出每一步的逻辑依据、可能的薄弱环节以及改进空间。这一做法将“CoT 可验证推理”从“隐性能力”

转化为“显性训练”。验证报告可设计为结构化模板，包含“每一步的依据是什么”“哪一步我最不确定”“AI 会如何评价这一步”等引导性问题。

第四，培养学生的“AI 批判性使用”意识。根据本文的调查结果，学生普遍缺乏判断 AI 输出正确性的能力(题项 Q8 的均值仅 3.76)，且这一问题在使用频率更高的学生中同样存在。建议在课程中专门设置“AI 输出批判性分析”环节，给学生提供若干份 AI 生成的证明(其中包含不同类型和程度的错误)，要求学生逐条指出错误并给出修正方案。这一训练将有效提升学生对 AI 输出的“免疫力”。具体可设计为“找茬式”课堂练习：每份 AI 生成的证明中包含 3~5 处逻辑错误，学生需要在限定时间内找出并修正，教师随后讲解每处错误的本质。

第五，逐步撤除 AI 辅助，促进能力内化。调查结果中，题项 Q7 (独立完成证明，M = 3.90)的得分偏低，说明学生在脱离 AI 辅助后的独立能力仍需加强。建议在教学过程中采用“渐进式撤除”策略：初期允许学生充分使用 AI 辅助，中期限制 AI 的使用场景，后期要求学生在无 AI 辅助的情况下完成限时证明。这一过程旨在将 CoT 推理从“外部工具”逐步内化为学生的“思维习惯”。具体时间安排可参考：前 4 周充分开放，中间 4 周限制使用场景，最后 2 周完全禁用，以此模拟“脚手架逐步撤除”的教学逻辑。

第六，将 CoT 可验证推理能力纳入课程考核体系。建议在期末考试中设置专门的“推理链验证”题型，给学生一个包含若干逻辑步骤的 CoT 推理链(其中混有正确步骤和错误步骤)，要求学生逐一判断每一步的正确性并说明理由。这种题型直接考察学生的 CoT 可验证推理能力，与传统的“构造证明”题型形成互补，能够更全面地评估学生的推理素养。

6. 结论与展望

本研究通过对 29 名离散数学学生的问卷调查发现：AI 工具已深度嵌入学生的学习过程，学生在“理解分步推理”和“清晰表达思路”方面获益明显，但在“独立判断 AI 生成证明的正确性”这一核心能力上存在显著短板。学生正在成为 CoT 推理链的“消费者”，却尚未成为 CoT 推理链的“验证者”。这一能力落差对离散数学教学提出了挑战，传统的“教师讲、学生练”模式已无法应对 AI 时代的新格局，教学亟需从“答案导向”转向“CoT 可验证推理导向”。

本研究也存在一定局限性。首先，样本量较小(29 人)，且全部来自同一班级，研究结论的推广性有待进一步验证。未来教学改革措施实施过程中，研究应显著扩大样本量，并覆盖不同类型高校和专业的学生，以提升研究普遍性。其次，本研究采用的是学生自评量表，缺乏客观的推理能力测试数据作为佐证。自我报告能够有效反映学生的主观感知与态度，但其与真实能力之间可能存在系统性偏差。未来研究可在此基础上，设计并使用客观的能力评测工具(例如，包含逻辑陷阱的 AI 生成证明、CoT 推理链，要求学生找出错误)，将学生的自我评价与客观表现进行对比分析，以更精确地刻画学生的 CoT 可验证推理能力现状。此外，未来研究还可开展准实验研究，设立实验班(采用文中所提的四阶段教学框架)和对照班(采用传统教学模式)，通过前测和后测比较两组学生在 CoT 可验证推理能力上的差异，以验证该教学框架的实际效果。

总之，一个基本判断已经清晰：在推理智能体能够生成 CoT 推理链的时代，离散数学教学的核心任务不再是“教会学生模仿证明”，而是“教会学生验证证明”。当 AI 能够“思考”时，教会和引导学生正确批判性地使用 AI 工具，或许是 AI 时代离散数学教育教学最紧迫要解决的问题。

基金项目

杭州电子科技大学高等教育教学改革研究项目“推理智能体重构离散数学 Chain-of-Thought 可验证

推理教学模式研究”(YBJG202615)。

参考文献

[1] 梁苗苗, 喻玲娟, 王慧, 等. 面向人工智能专业的“离散数学”教学改革[J]. 萍乡学院学报, 2025, 42(3): 90-94.

[2] Trinh, T.H., Wu, Y.H., Le, Q.V., He, H. and Luong, T. (2024) Solving Olympiad Geometry without Human Demonstrations. *Nature*, **625**, 476-482. https://doi.org/10.1038/s41586-023-06747-5

[3] Ringer, T. (2025) Mathematicians Put AI Model AlphaProof to the Test. *Nature*, **651**, 595-597. https://doi.org/10.1038/d41586-025-03585-5

[4] Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., *et al*. (2025) DeepSeek-R1 Incentivizes Reasoning in LLMs through Reinforcement Learning. *Nature*, **645**, 633-638. https://doi.org/10.1038/s41586-025-09422-z

[5] 张妙雨, 张琼, 周芃杜. 从“证明者”到“鉴赏者”: 第五范式下数学科研人才培养的认知转向[EB/OL]. https://chinaxiv.org/abs/202512.00303, 2025-11-27.

[6] Hanna, G., Larvor, B. and Yan, X.K. (2024) Using the Proof Assistant Lean in Undergraduate Mathematics Classrooms. *ZDM—Mathematics Education*, **56**, 1517-1529. https://doi.org/10.1007/s11858-024-01577-9

[7] Stanford CS109 Staff (2024) CS109: Probability for Computer Scientists—Course Materials. Stanford University.

[8] Kawski, M. (2025) ChatGPT and AI Enhancing Undergrad Math. 2025 *Proceedings of the Asian Technology Conference in Mathematics*, Manila, 13-16 December 2025, 1.

[9] 谢鑫, 赵正. 课程思政背景下基于知识图谱和多专家智能体的离散数学教学探索[J]. 计算机教育, 2025(10): 191-195.

[10] 李琳. 实践教学在离散数学教改中的探索[J]. 物流工程与管理, 2025, 47(2): 174-176.

[11] Likert, R. (1932) A Technique for the Measurement of Attitudes. *Archives of Psychology*, **22**, 140-155.

[12] Cheng, Z.H., Dai, W., Sun, J.H., *et al*. (2026) Imbuing Large Language Models with Bidirectional Logic for Robust Chain Repair. https://arxiv.org/abs/2606.05030

[13] Shao, J.T. and Cheng, Y.M. (2025) Cot Is Not True Reasoning, It Is Just a Tight Constraint to Imitate: A Theory Perspective. https://arxiv.org/abs/2506.02878

[14] You, W., Xue, A., Havaldar, S., Rao, D., Jin, H., Callison-Burch, C., *et al*. (2025) Probabilistic Soundness Guarantees in LLM Reasoning Chains. *Proceedings of the* 2025 *Conference on Empirical Methods in Natural Language Processing*, Suzhou, November 2025, 7517-7536. https://doi.org/10.18653/v1/2025.emnlp-main.382

[15] Sweller, J. (1988) Cognitive Load during Problem Solving: Effects on Learning. *Cognitive Science*, **12**, 257-285. https://doi.org/10.1207/s15516709cog1202_4

[16] 赖日生, 曾晓青, 陈美荣. 从认知负荷理论看教学设计[J]. 南昌师范学院学报, 2005, 26(1): 52-55.

[17] Wood, D., Bruner, J.S. and Ross, G. (1976) The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, **17**, 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

[18] van de Pol, J., Volman, M. and Beishuizen, J. (2010) Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review*, **22**, 271-296. https://doi.org/10.1007/s10648-010-9127-6

[19] 王颖. 维果茨基最近发展区理论及其应用研究[J]. 山东社会科学, 2013(12): 180-183.

[20] 王嘉欣. 认知卸载与社会性线索对元认知监控的影响[D]: [硕士学位论文]. 金华: 浙江师范大学, 2023.

[21] Tomisu, H., Ueda, J. and Yamanaka, T. (2025) The Cognitive Mirror: A Framework for AI-Powered Metacognition and Self-Regulated Learning. *Frontiers in Education*, **10**, Article 1697554. https://doi.org/10.3389/feduc.2025.1697554

[22] Curran, D., Chapman, B. and Conway, M. (2025) Utilizing LLMs to Investigate the Disputed Role of Evidence in Electronic Cigarette Health Policy Formation in Australia and the UK. https://doi.org/10.48550/arXiv.2505.06782

[23] Lipponen, L. (2023) Exploring Foundations for Computer-Supported Collaborative Learning. In: *Computer Support for Collaborative Learning*, Routledge, 72-81. https://doi.org/10.4324/9781315045467-12