

人工智能作为《热学》课程答疑工具的测试和探讨

蓝善权

岭南师范学院物理科学与技术学院, 广东 湛江

收稿日期: 2025年5月14日; 录用日期: 2025年7月1日; 发布日期: 2025年7月10日

摘要

本文探讨了利用现成人工智能模型作为《热学》课程答疑工具的可行性。通过对DeepSeek-V3、DeepSeek-R1、ChatGPT-4.1 mini、Claude-3.5 Haiku和Gemini-2.5 Flash五款模型进行测试, 发现DeepSeek-R1表现最佳, 能够胜任辅助答疑工作; Gemini-2.5 Flash、ChatGPT-4.1 mini和DeepSeek-V3也可有效减轻教师负担; 而Claude-3.5 Haiku则不建议作为答疑工具。测试结果表明, 人工智能模型在解答概念题方面表现优秀, 在解答进阶性填空题、单选题和计算题方面也展现出一定的能力。在人工智能时代, 教师培养学生提出问题的能力比以往显得更加重要。本文的研究可以推广到其他课程, 对人工智能辅助教学具有重要的参考价值。

关键词

课程答疑, 人工智能, 热学, 可行性测试

Test and Discussion on Artificial Intelligence as a Q&A Tool for the “Thermodynamics” Course

Shanquan Lan

School of Physics Science and Technology, Lingnan Normal University, Zhanjiang Guangdong

Received: May 14th, 2025; accepted: Jul. 1st, 2025; published: Jul. 10th, 2025

Abstract

This paper explores the feasibility of using existing artificial intelligence (AI) models as a Q&A tool

for the “Thermodynamics” course. By testing five models: DeepSeek-V3, DeepSeek-R1, ChatGPT-4.1 mini, Claude-3.5 Haiku, and Gemini-2.5 Flash, it was found that DeepSeek-R1 performed the best and is capable of assisting with Q&A tasks. Gemini-2.5 Flash, ChatGPT-4.1 mini and DeepSeek-V3 can also effectively reduce the burden on instructors, while Claude-3.5 Haiku is not recommended as a Q&A tool. The test results indicate that AI models excel in answering conceptual questions and demonstrate a certain level of ability in addressing advanced fill-in-the-blank questions, multiple-choice questions, and calculation problems. In the era of artificial intelligence, cultivating students’ ability to ask questions is more important than ever for teachers. This research can be generalized to other curricula and provides valuable insights for AI-assisted teaching.

Keywords

Course Q&A, Artificial Intelligence, Thermodynamics, Feasibility Test

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

《热学》是我校物理学和应用物理学专业的核心专业课程。这两个专业 2024 级的学生近 400 人，学生学习主动性差异大，教师答疑压力大。因此利用人工智能作为辅助答疑工具很有必要，此外它还具有诸多优势，比如随时随地答疑、给学生轻松问答环境、个性化学习支持等优点。总之，教师答疑和人工智能辅助答疑相结合，一方面可以减轻教师负担，另一方面将给学生更多的答疑途径。

随着人工智能技术的快速进步，其在教育领域的应用逐渐成为全球关注的焦点。人工智能通过大数据分析、自然语言处理、机器学习等技术，为教育带来了个性化学习、智能评估、资源优化、科学管理等创新，同时推动了教育公平和教育效率的提升[1][2]。人工智能作为课程答疑工具目前主流的方法是建立一个基于人工智能的答疑系统[3][4]，即首先创建课程的数据库，然后把所有课程资料提供给人工智能学习并调试。显然这种方法需要一定的编程能力和资金支持。那有没有更简单的方法呢？随着人工智能模型的不断优化更新，其智力越来越高，现行的人工智能是否可以直接作为答疑工具呢？这是一个重要的具有现实教学意义的研究课题，如果可行，将极大地帮助广大教育工作者在教学中应用人工智能。本文将通过《热学》课程中的问题测试验证该方法的可行性。

2. 实验设计和测试

人工智能作为课程的辅助答疑工具需要满足三个条件：成本低、响应快、回答准确。由此我们选定了五款人工智能模型作为测试对象，分别是 DeepSeek-V3 (2024 年 12 月 26 日发布)、DeepSeek-R1 (2025 年 1 月 20 日发布)、ChatGPT-4.1 mini (2025 年 4 月 15 日发布)、Claude-3.5 Haiku (2024 年 10 月 22 日发布)和 Gemini-2.5 Flash (2025 年 4 月 10 日发布)。它们都具有低延迟的特点，基本可以实现实时交互和免费，详情可以查看文献[5][6]。

测试内容为《热学》课程中的重难点知识，比如热力学第零定律、第一定律、第二定律、第三定律、气体分子动理论、气体分子速率和能量分布律、气体输运过程等。问题集(参看附录)包含 10 个概念题、5 个填空题、5 个单选题和 5 个计算题，其中概念题属于基础性题目(30%)，填空、单选、计算题属于进阶性题目(70%)。问题集里的 25 个问题是来自课本和题库中精心挑选的，它们涵盖了我校《热学》课程教

学的各章节的内容，具有代表性。为了检验题目的难易程度，在人工智能模型测试之前，我们把问题集制作成试卷类型的 Word 文件，发放给一个班级的学生进行测试。该班级共 51 位同学，成绩统计如表 1 所示。从中可以发现各题型的得分率相差不大，而概念题和计算题的得分率比填空题和单选题稍微高一些；总平均分 70.1 分，试题难度适中。总而言之，试题的知识点覆盖面广，难易程度适中，能充分反映出被测试者对《热学》课程的掌握程度。

Table 1. Statistics of test scores for students (51 participants)

表 1. 学生(51 位)做测试题的成绩统计

题型	概念题(30 分)	填空题(15 分)	单选题(15 分)	计算题(40 分)
平均分	22.06	10.12	9.47	28.45
总平均分	70.1			

人工智能模型的测试过程：把 word 版的试题上传给各个人工智能模型，输入简单指令“这是一份热学试卷，请做完，要求写出重要的步骤”；下载人工智能的答卷，对照参考答案批阅。五款人工智能模型做完所有试题用时都不超过 5 分钟，完全符合答疑实时交互的要求。按照解题的三步骤进行错误类型分析：(1) 读题性错误，即没有理解题意，或知识理解不到位；(2) 逻辑性错误，即没有解题思路，或者解题思路不对；(3) 计算性错误，即数学运算出错。表 2 是人工智能做测试题的结果统计。

Table 2. Statistics of test results for artificial intelligence

表 2. 人工智能做测试题的结果统计

人工智能模型		DeepSeek-V3	DeepSeek-R1	ChatGPT-4.1 mini	Claude-3.5 Haiku	Gemini-2.5 Flash
概念题(每小题 3 分, 共 30 分)	得分	30	30	30	30	30
	错题	无	无	无	无	无
填空题(每小题 3 分, 共 15 分)	得分	12	15	9	6	12
	错题	第 4 题。	无	第 3 题, 第 4 题。	第 1 题, 第 3 题, 第 4 题。	第 3 题。
单选题(每小题 3 分, 共 15 分)	得分	15	15	9	9	12
	错题	无	无	第 3 题, 第 4 题。	第 3 题, 第 4 题。	第 4 题。
	得分	30	40	40	18	40
计算题(每小题 8 分, 共 40 分)	得分	30	40	40	18	40
	错题	第 2 题(扣 4 分), 第 3 题(扣 6 分)。	无	无	第 1 题(扣 2 分), 第 2 题(扣 2 分), 第 3 题(扣 8 分), 第 4 题(扣 4 分), 第 5 题(扣 6 分)。	无
总分		87	100	88	63	94

DeepSeek-V3 的错题案例。(1) 填空题第 4 题：在利用热力学第一定律计算系统对外界做功的时候正负号反了，即系统对外界做功和外界对系统做功分不清楚，属于逻辑性错误。(2) 计算题第 2 题：两问的步骤都对，但是计算结果错误，属于计算性错误。(3) 计算题第 3 题：第一步是对的，但是后续解题思路错误，属于逻辑性错误。总的来说，DeepSeek-V3 在对问题的理解方面表现很优秀，理解问题很到位；在解决问题的逻辑思维和能力方面表现良好，能解决大部分的问题；在计算能力方面表现良好，能计算正确大部分的结果，但是计算题的第 2 题的运算只涉及到乘除和开方，它居然计算错误。

DeepSeek-R1 没有错题。DeepSeek-R1 在理解题意、解题思路和计算能力方面都表现优秀。

ChatGPT-4.1 mini 的错题案例。(1) 填空题第 3 题：刚性多原子分子的转动自由度为 3，双原子分子

的转动自由度为 2, 单原子分子转动自由度为 0, 解题过程中错误的把氧气分子转动自由度当成 3, 而非 2, 属于读题性错误。(2) 填空题第 4 题: 同 DeepSeek-V3。(3) 单选题第 3 题: 错误的选择 D, 对熵增加原理的理解不全面, 属于读题性错误。(4) 单选题第 4 题: 错误的选择 A, 对热力学第二定律的开尔文表述理解不全面, 属于读题性错误。总的来说, ChatGPT-4.1 mini 在对问题的理解方面表现良好, 基本到位; 在解决问题的逻辑思维和能力方面表现良好, 能解决大部分的问题; 在计算能力方面表现很优秀, 全部计算正确。

Claude-3.5 Haiku 的错题案例。(1) 填空题第 1 题: 乘法和幂指数计算错误, 属于计算错误。(2) 填空题第 3 题: 错误的把氧气分子转动自由度当成 0, 而非 2, 属于读题性错误。(3) 填空题第 4 题: 同 DeepSeek-V3。(4) 单选题第 3 题: 同 ChatGPT-4.1 mini。(5) 单选题第 4 题: 同 ChatGPT-4.1 mini。(6) 计算题第 1 题: 最后一步涉及的减乘除运算错误, 属于计算性错误。(7) 计算题第 2 题: 第一问方均根速率并未计算出结果, 属于计算性错误。(8) 计算题第 3 题: 答非所问, 属于读题性错误。(9) 计算题第 4 题: 第一问和第三问涉及的除乘减法运算错误, 属于计算性错误。(10) 计算题第 5 题: 只完成了第一步, 没有后续作答, 可能是没有理解题意或者没有解题思路, 属于读题性错误或者逻辑性错误。总的来说, Claude-3.5 Haiku 在理解题意、解题思路方面表现一般, 在计算能力方面表现不足。

Gemini-2.5 Flash 的错题案例。(1) 填空题第 3 题: 错误的把氧气分子的振动自由度当成了转动自由度, 属于读题性错误。(2) 单选题第 4 题: 同 ChatGPT-4.1 mini。总的来说, Gemini-2.5 Flash 在对问题的理解方面表现良好, 基本到位; 在解题思路和计算能力方面表现优秀。

通过试题的测试, 从表 1 的统计可以发现, (1) 测试成绩 DeepSeek-R1 > Gemini-2.5 Flash > ChatGPT-4.1 mini > DeepSeek-V3 > Claude-3.5 Haiku。其中 DeepSeek-R1 满分, 完全可以胜任《热学》课程的辅助答疑工作。Gemini-2.5 Flash 94 分, 可以较好胜任辅助答疑工作。ChatGPT-4.1 mini 和 DeepSeek-V3 接近 90 分, 错误较少, 可以作为辅助答疑工具, 减轻教师答疑负担。而 Claude-3.5 Haiku 刚刚及格, 不建议作为辅助答疑工具。(2) 对基础性的概念题, 五个人工智能模型都回答的准确无误, 可以用来帮助学生理解物理概念和规律。对进阶性的填空题和单选题, DeepSeek 系列表现优秀, 它们的中文理解很准确, 而 Gemini-2.5 Flash 表现良好, ChatGPT-4.1 mini 和 Claude-3.5 Haiku 表现一般, 可能是中文资料库不全, 它们的中文理解能力不如 DeepSeek 系列[7]。对进阶性的计算题, DeepSeek-R1、ChatGPT-4.1 mini 和 Gemini-2.5 Flash 满分, 表现优秀; DeepSeek-V3 存在逻辑性和计算性错误, 表现一般; Claude-3.5 Haiku 存在读题性、逻辑性和计算性错误, 表现不及格。(3) 同一系列 DeepSeek-R1 和 DeepSeek-V3 的专注领域有所不同, DeepSeek-R1 更专注于逻辑推理和学术, 更适合作为《热学》课程的答疑工具。

以上模型在答题表现上的差异源于模型的结构、参数、训练数据、应用场景等因素。DeepSeek-V3 作为 DeepSeek 系列模型的重要组成部分, 采用 MoE 架构, 总参数量达到 671B, 专注于多语言翻译、图像生成和 AI 绘画等任务。DeepSeek-R1 是一款轻量级推理模型, 参数量仅为 DeepSeek-V3 的十分之一, 但具备强大的推理能力, 支持多模态处理, 适用于学术研究和决策支持系统。ChatGPT-4.1 mini 是 OpenAI 推出的 GPT-4.1 系列模型之一, 专注于实时交互和轻量级任务, 具有低延迟和高性价比的特点, 在代码生成和编程任务方面具有强大能力。Claude-3.5 Haiku 由 Anthropic 公司发布, 专注于低延迟和精准工具使用能力, 适用于用户界面产品和大规模数据处理。Gemini 2.5 Flash 由谷歌发布, 专注于推理效率, 特别适用于高容量和实时应用场景, 如客户服务和文档解析。总之, DeepSeek-R1 表现出色是因为它强大的推理能力和中文理解能力。

3. 总结与讨论

为了研究人工智能模型直接作为《热学》课程的答疑工具的可行性, 我们首先收集了《热学》课程

中的重难点知识,制作了一份包含概念题、填空题、单选题和计算题的试卷,该试题的知识点覆盖面广、难度适中。然后用这份试卷对 DeepSeek-V3、DeepSeek-R1、ChatGPT-4.1 mini、Claude-3.5 Haiku 和 Gemini-2.5 Flash 五个人工智能模型进行测试。最后通过测试的统计结果分析发现直接把 DeepSeek-R1、Gemini-2.5 Flash、ChatGPT-4.1 mini 和 DeepSeek-V3 作为《热学》课程的辅助答疑工具是可行的,其中 DeepSeek-R1 表现最优,而不建议把 Claude-3.5 Haiku 作为答疑工具。

我们的研究可以推广到各类课程,特别是类似《热学》的理科课程,所以本文对人工智能辅助教学具有重要的参考价值。研究的创新表现在三个方面,(1) 研究内容创新,探究人工智能模型直接作为《热学》课程的答疑工具的可行性;(2) 横向(DeepSeek、ChatGPT、Claude、Gemini)和纵向(DeepSeek-V3、DeepSeek-R1)全面地比较目前热门的几个人工智能模型;(3) 创新性地根据解题的三步骤将错题类型归纳为读题性错误、逻辑性错误和计算性错误。然而,研究尚存在一点不足,即缺少人工智能作为答疑工具的教学实践的反馈,这需要对学生的使用情况和效果进行调研,它将作为未来的一个研究内容。同时如何预防学生利用人工智能做作业也是当下一个重要的研究课题。随着人工智能模型的不断优化更新(比如即将发布的 DeepSeek-R2),我们相信未来将有更多的人工智能可以胜任各类课程的辅助答疑工作,而不需要花费大量资金用于建立专用的辅助答疑智能体。现在的教师既要培养学生解决问题的能力,更要培养学生提出问题的能力,而可以把部分问题的求解交给人工智能。

基金项目

人工智能赋能《热学》课程教学的探索。岭南师范学院教改项目,2024。

参考文献

- [1] 余亮, 邓双洁, 张馨月. 人工智能技术赋能教育的演进脉络、内在逻辑和发展趋势[J]. 电化教育研究, 2025, 46(6): 13-20+28.
- [2] 张蕾, 郭茂祖. 人工智能会替代人类教师吗?[J]. 教育进展, 2022, 12(1): 6-9.
- [3] 韩英杰, 张坤丽. 高校人工智能自动答疑系统构建研究及应用[J]. 信息与电脑(理论版), 2024, 36(2): 132-135.
- [4] 谢珺, 杨海洋, 梁凤梅, 等. 基于课程图谱的智能答疑系统设计与开发——以“信号与系统”为例[J]. 系统科学学报, 2025(3): 161-167.
- [5] 朱嘉明. 人工智能进化尺度和大模型生态——DeepSeek V3 和 R1 系列现象解析[J]. 经济导刊, 2025(2): 36-45.
- [6] 李嘉慧. 超越传统边界: 生成式 AI (GAI)在教育领域的颠覆性创新与协同共生[J]. 汉字文化, 2025(8): 41-43.
- [7] 伍革新, 涂乐, 邱帆. 大模型环境下的中英文语言对比分析[J]. 今古文创, 2025(14): 134-136.

附录

测试题目和答案

一、概念题(每小题 3 分, 共 30 分)

1、热力学第零定律。

如果两个系统分别与第三个系统处于热平衡, 那么这两个系统也处于热平衡。

2、理想气体物态方程的表达式。

$$pV = nRT$$

3、水的三相点的温度值(单位用开尔文)。

273.16 K。

4、能量按自由度均分定理。

在温度为 T 的平衡状态下, 物质分子的每一个自由度都具有相同的平均动能, 其大小等于 $\frac{1}{2}kT$ 。

5、理想气体平均自由程的数学表达式。

$$\lambda = \frac{1}{\sqrt{2}\pi d^2 n} \text{ 或 } \lambda = \frac{kT}{\sqrt{2}\pi d^2 p}$$

6、热力学第一定律。

系统从外界吸收的热量等于系统内能的增加量与系统对外做的功之和, $Q = \Delta U + W$ 。

7、热力学第二定律。

克劳修斯表述: 热量不能自发地从低温物体传递到高温物体。

开尔文表述: 无法从单一热源取热并完全转化为有用功而不产生其他影响。

8、卡诺定理。

可逆热机的效率是所有工作在同一温差的热机中最高的。

9、熵增原理。

孤立系统的熵永不减少。

10、热力学第三定律。

当温度接近绝对零度时, 系统的熵趋近于零。

或者: 不可能在有限的过程下使一个物体冷却到绝对温度的零度。

二、填空题(每小题 3 分, 共 15 分)

1、一容器内装有理想气体, 其真空度, 即气体压强为 10^{-5} mmHg, 则温度为 300 K 时管内单位体积的分子数为(3.22×10^{17})。

2、给定的理想气体, 比热容比 γ 为已知, 从初态 p_0 、 V_0 、 T_0 开始, 作绝热膨胀, 体积增大到三倍, 膨胀后的温度 $T = (3^{1-\gamma} T_0)$ 。

3、在温度为 400 K 时, 1 mol 氧气分子的转动动能为(3325.6) J, 其分子可视为刚性分子。

4、一气缸内储有 10 mol 的刚性单原子分子理想气体, 在压缩过程中外界做功 209 J, 气体升温 1 K, 此过程中外界传给气体的热量为(-84.3) J。

5、在温度分别为 600 K 和 300 K 的高温热源和低温热源之间工作的热机, 理论上的最大效率为(50%)。

三、单选题(每小题 3 分, 共 15 分)

1、一定量的理想气体, 在体积不变的条件下, 当温度升高时, 分子的平均碰撞频率和平均自由程的变化情况是(A)。

A. 平均碰撞频率增大, 平均自由程不变; B. 平均碰撞频率不变, 平均自由程增大;

C. 平均碰撞频率和平均自由程都增大; D. 平均碰撞频率和平均自由程都不变。

2、有人设计了一台可逆的卡诺热机, 每循环一次可从 400 K 的高温热源吸热 1800 J, 向 300 K 的低温热源放热 800 J, 同时对外做功 1000 J, 这样的设计是(D)。

- A. 可以的, 符合热力学第一定律;
- B. 可以的, 符合热力学第二定律;
- C. 不行的, 卡诺循环所做的功不能大于向低温热源放出的热量;
- D. 不行的, 这个热机的效率超过了理论值。

3、关于可逆过程和不可逆过程的判断正确的是(A)

- A. 可逆热力学过程一定是准静态过程;
- B. 准静态过程一定是可逆过程;
- C. 不可逆过程就是不能向相反方向进行的过程;
- D. 熵增加的过程是不可逆过程。

4、热力学第二定律表明(C)。

- A. 不可能从单一热源吸收热量使之全部变为有用功;
- B. 在一个可逆过程中, 工作物质净吸热等于对外做的功;
- C. 摩擦生热的过程是不可逆的;
- D. 热量不可能从温度低的物体传到温度高的物体。

5、速率分布函数 $f(v)$ 的物理意义为(B)。

- A. 具有速率 v 的分子占总分子数的百分比;
- B. 速率分布在 v 附近的单位速率间隔中的分子数占总分子数的百分比;
- C. 具有速率 v 的分子数;
- D. 速率分布在 v 附近的单位速率间隔中的分子数。

四、计算题, 需要写出计算过程(每小题 8 分, 共 40 分)

1、一氧气瓶的容积是 32 L, 其中氧气的压强是 130 atm, 规定瓶内氧气压强降到 10 atm 时就得充气, 以免混入其他气体而需洗瓶, 今有一玻璃室, 每天需用 1.0 atm 氧气 400 L, 问一瓶氧气能用几天? (假设氧气可视理想气体, 且在使用氧气过程中温度 T 不变)

答案: 9.6 天。

2、气体的温度为 $T = 273 \text{ K}$, 压强为 $p = 1.00 \times 10^{-2} \text{ atm}$, 密度为 $\rho = 1.29 \times 10^{-5} \text{ g/cm}^3$

(i) 求气体分子的方均根速率。

(ii) 求气体的分子量, 并确定它是什么气体。

答案: (i) 485 m/s; (ii) 28.9, 空气。

3、0.0080 kg 氧气, 原来温度为 300 K, 体积为 0.41 L, 若: 经过绝热膨胀体积增加至 4.1 L; 试计算外界对气体所做的功, 设氧气可看作刚性理想气体。

答案: -937.5 J。

4、1 mol 单原子理想气体在容积保持不变的情况下从 300 K 加热到 400 K, 问增加了多少内能? 对外做了多少功? 吸收了多少热量?

答案: 1247.1 J; 0 J; 1247.1 J。

5、一可逆卡诺热机低温热源的温度为 280 K, 效率为 40%。若要将其效率提高到 50%, 则高温热源的需提高多少开尔文?

答案: 93.3 K。