

生成式AI在“数据分析与挖掘”课程资源构建中的应用研究

赵 猛, 王双明, 伍建全, 刘金源, 邹细涛

重庆科技大学计算机科学与工程学院, 重庆

收稿日期: 2024年11月26日; 录用日期: 2024年12月24日; 发布日期: 2024年12月31日

摘 要

生成式AI是一种借助现代深度学习技术的文本(包括代码)、图像和视频生成工具。数据分析与挖掘课程以其技术复杂、工具抽象、代码繁多为特点,需要对众多python工具命令完全掌握才能够比较熟练地上手操作。此外,该课程的学习资源却较为分散,案例代码来源不够清楚,代码注解不够详细,从而使得课程资源的构建需要花费较多时间。结合最新的生成式AI技术(本文使用文心一言、Gamma、Github Copilot 3种工具),教师可以直接通过聊天或提示等方式输入文本提示词,进而采用生成式AI工具直接输出教案、PPT和代码等课程资源,教师只需要负责审核生成结果的有效性和正确性即可,极大地提高了工作效率和准确性,为人工智能在专业型课程资源构建方式方法中的应用提供了借鉴。

关键词

生成式AI, 数据分析与挖掘, 资源构建, 应用研究

Research on Application of Generative AI in the Resource Construction of “Data Analysis and Mining” Course

Meng Zhao, Shuangming Wang, Jianquan Wu, Jinyuan Liu, Xitao Zou

School of Computer Science and Technology, Chongqing University of Science and Technology, Chongqing

Received: Nov. 26th, 2024; accepted: Dec. 24th, 2024; published: Dec. 31st, 2024

Abstract

Generative AI is a tool for text (including code), image and video generation with the help of modern deep learning techniques. The Data Analysis and Mining course is characterized by its technical

文章引用: 赵猛, 王双明, 伍建全, 刘金源, 邹细涛. 生成式AI在“数据分析与挖掘”课程资源构建中的应用研究[J]. 教育进展, 2024, 14(12): 1441-1449. DOI: 10.12677/ae.2024.14122435

complexity, tool abstraction, and extensive code, which requires a complete mastery of numerous python tool commands to be able to operate it more proficiently. In addition, the learning resources of the course are scattered, the source of the case code is not clear enough, and the code annotations are not detailed enough, thus making the construction of the course resources take more time. Combined with the latest generative AI technology (in this paper, we use three tools: Wenxin Yiyin, Gamma, and Github Copilot), teachers can directly enter text prompts through chat or prompts, and then use generative AI tools to directly output course resources, such as lesson plans, PPTs, and code, etc., and the teacher only needs to be responsible for reviewing the validity and correctness of the generated results, which greatly improves the efficiency and accuracy of the work, and provides a reference for the application of AI in the construction of professional course resources.

Keywords

Generative AI, Data Analysis and Mining, Resource Construction, Application Research

Copyright © 2024 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

“数据分析与挖掘”课程总学时数为 48 学时左右,要求学生能够利用 Numpy、Pandas、Keras 等 Python 工具,完成对数据集的加载、预处理(空值、重复值、异常值处理等)、挖掘和可视化,其中挖掘部分涉及具体算法的实现,主要包括:分类模型、回归分析、聚类、时序分析等内容。“数据分析与挖掘”课程是计算机科学与技术、智能科学与技术、数据科学与大数据技术等专业的专业基础课,是数据分析理论与实践应用相结合的课程,它担负着使学生系统且全面地理解数据的价值,并应用生产生活中的各种数据创造社会价值的使命。

生成式 AI 作为人工智能领域的一大分支[1],近年来备受关注。其以文本、图像、视频等生成成为特色,以大数据为背景,依托深度学习大模型进行广泛训练,并通过特殊泛化手段,得到相应提示词下的全新、逼真的文本、图像、视频内容等结果。改变了传统的人工模式,极大地提高了生产效率,解放了人力。生成式 AI 的核心在于其能够从大量数据中学习并模拟出数据的内在规律和分布特征,进而生成与真实数据相似但又全新的内容。在技术上,生成式 AI 的发展经历了多个阶段。早期的生成式 AI 主要基于传统的统计学习方法,如隐马尔可夫模型、朴素贝叶斯等,这些方法在一定程度上能够实现数据的生成,但受限于模型的复杂度和计算能力,生成的效果往往不尽如人意[2]。随着深度学习的兴起,尤其是神经网络技术的快速发展,生成式 AI 迎来了新的突破。目前,主流的生成式 AI 技术包括生成对抗网络(GAN)、变分自编码器(VAE)、自回归模型(如 GPT 系列)等,这些技术在图像生成、文本生成、语音合成等领域取得了令人瞩目的成果[3]。

2. “数据分析与挖掘”课程资源构建中存在的问题

在课程资源构建方面,生成式 AI 同样具有广阔的应用前景[2][3]。传统的课程资源构建往往依赖于人工编辑和整理,不仅效率低下,而且难以保证课程资源的更新和质量。而生成式 AI 则能够根据学生的学习需求和兴趣点,自动地生成个性化的学习资源和推荐路径,例如,通过分析学生的学习数据和行为模式[4][5]。生成式 AI 可以为学生量身定制学习计划,推荐适合其学习风格和进度的课程资源;同时,生成式 AI 还可以用于智能答疑、在线评估等场景,提升学生的学习体验和效果。综上所述,生成式 AI

在数据分析与挖掘课程资源构建中的应用具有重要的研究价值和实际意义[3]。通过深入研究生成式 AI 的相关技术和方法,结合具体的应用场景和需求,我们可以进一步挖掘数据的价值,提升课程资源的构建效率和质量,为教育领域的发展注入新的活力。

随着信息技术的迅猛发展,数据分析与挖掘已成为多个领域不可或缺的技能。然而,在数据分析与挖掘课程资源的构建方面,我们仍面临诸多挑战。本文旨在深入探讨这些挑战,并分析当前形势下对本门课程资源的需求,以期对相关教育工作者提供有益的参考[4]。

首先是技术挑战。数据分析与挖掘涉及众多复杂的技术和算法,如何将这些技术以易于理解的方式呈现给学生,是一个亟待解决的问题。其次是内容挑战。由于数据分析与挖掘的应用领域广泛,课程内容的选择和安排需要充分考虑不同领域的需求和特点。最后是教学挑战。如何激发学员的学习兴趣,提高他们的实践操作能力,是课程构建过程中必须关注的重要问题[5]。

与此同时,行业内对数据分析与挖掘课程资源的需求也日益旺盛。随着大数据时代的来临,越来越多的企业和组织开始重视数据分析与挖掘在业务决策中的应用。他们急需具备相关技能的人才来支持业务的发展和创新。因此,对于高质量、实用性强的数据分析与挖掘课程资源的需求也日益迫切[3]。

综上所述,数据分析与挖掘课程资源的构建是一个复杂而重要的任务。我们需要正视当前的挑战,深入分析行业内的需求,并从多个方面入手来改进课程资源的构建。只有这样,我们才能培养出更多具备相关技能的人才,为行业的发展和创新提供有力的支持。

3. 生成式 AI 辅助“数据分析与挖掘”课程资源构建的途径

为了满足教学相关需求,我们需要从多个方面着手来改进数据分析与挖掘课程资源的构建方式。首先,我们应该注重理论与实践的结合。通过引入实际案例和项目,让学生在实践中掌握所学知识,提高他们的操作技能。其次,我们需要及时更新课程内容,确保学生能够接触到最前沿的知识和技术。此外,我们还可以通过多样化的教学方式和手段来激发学员的学习兴趣,提高他们的学习效果。而生成式 AI 可以通过提示词较为容易地根据要求生成各种教学材料,且更新及时。生成式 AI 在数据分析与挖掘课程资源构建中的理论基础,主要涉及到生成式 AI 与课程资源构建的结合点,以及生成式 AI 在课程资源构建领域的应用价值和与数据分析挖掘课程的契合度分析。

生成式 AI 与课程资源构建的结合点体现在多个方面:(1) 自动化内容生成——生成式 AI 具备强大的数据处理和生成能力,能够快速分析海量教育数据,提取有用信息,并生成高质量的教育资源;(2) 个性化学习策略——在课程资源构建中,生成式 AI 可以助力教师和学习者更有效地利用数据,采用个性化生成教学内容。例如,通过分析学生的兴趣点,生成式 AI 可以为学生推荐个性化的学习路径和资源,从而满足不同学生的需求。目前,已有一些生成式 AI 辅助教学的相关论文发表,但主要集中在教学过程的辅助,例如:文献[4]通过 CodeGeeX 工具生成课程教学代码;文献[6]通过生成式 AI 生成 C 语言教学代码;文献[7]-[9][12]探讨 ChatGPT 对话技术在课程中的应用;文献[10][11]探讨了教学模式和教学质量评估的人工智能辅助方法。生成式 AI 在课程资源构建领域的价值主要体现在:通过快速生成高质量的课程内容,在提升课程资源质量的同时减少教师的时间和精力;为学生提供个性化的学习资源和建议,满足他们的不同学习需求,提高学习效果。由于生成式 AI 当下的文本代码生成能力,而数据分析与挖掘课程存在大量以 Python 语言实现的算法代码,从而二者具有天然的契合度。

常见的生成式 AI 教学辅助工具如图 1 所示,其中主要罗列了课程资源辅助中的常见任务,包括:教案生成、PPT 生成和代码生成等主要功能。迄今,教案生成的主要工具包括:文心一言、通义千问、星火大模型、西沃教学大模型等;授课 PPT 生产工具主要包括:Gamma、Canva、文心一言、星火大模型等;代码生成和解释工具主要包括:Github Copilot、CodeGeeX、CodeWhisperer 等。从中可以看出,现

有的大模型通常不止一种功能，在各项任务中都展示了能力，例如：文心一言、星火大模型等等。

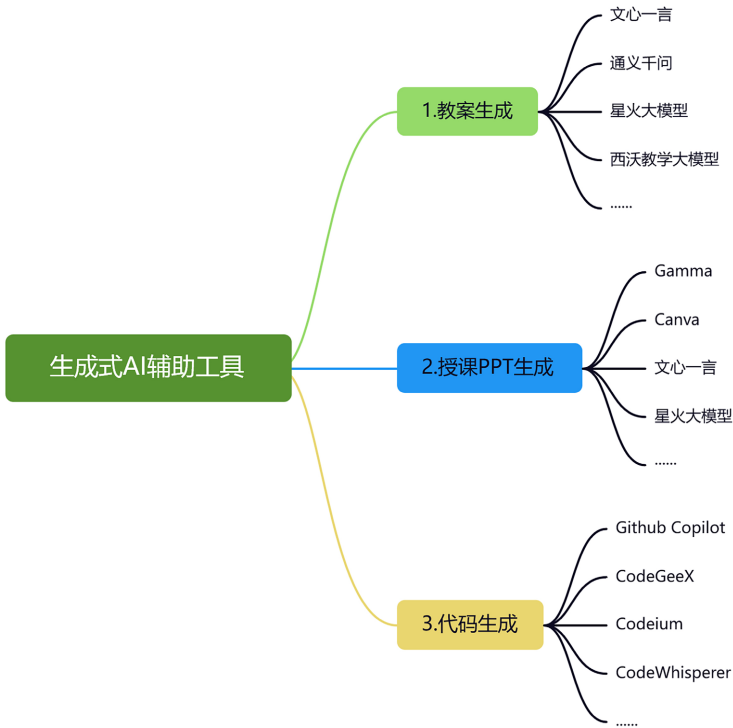


Figure 1. Generative AI tools for education
图 1. 生成式 AI 教育辅助工具

生成式 AI 辅助教育工具的使用步骤如图 2 所示。

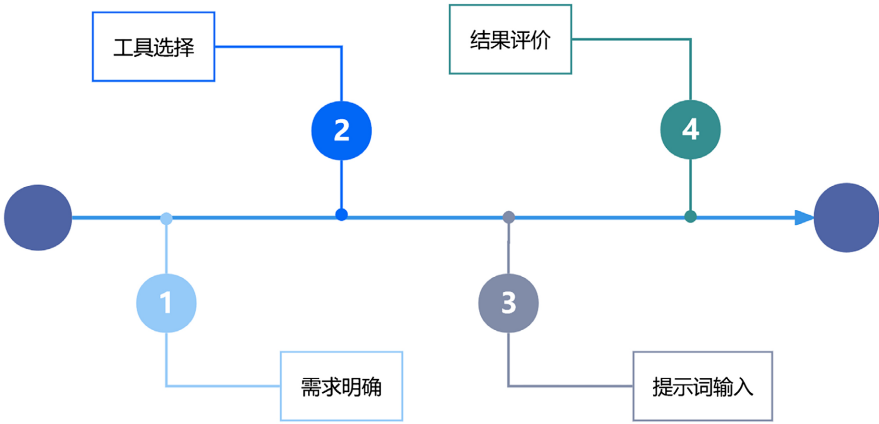


Figure 2. Steps for using Generative AI tools
图 2. 生成式 AI 工具使用步骤

阶段 1 (需求明确):

确定生成式 AI 所需生成的文档的类型，例如：PPT 课件、课程教案、课程代码等；确定生成的内容的主题，例如：python 代码生成中利用 pandas 去掉重复值等。

阶段 2 (工具选择): 根据任务类型，选择对应工具，例如：教案生成任务，可以参考图 1 所示选择文

心一言、通义千问等工具；授课 PPT 生成任务，可以选择 Gamma、Canva 等工具；代码生成任务，可以选择 Github Copilot、CodeGeeX 等工具。

阶段 3 (输入提示词): 根据生成内容的主题，采用自然语言描述需求，要求尽量避免模糊和抽象的词汇，指令清晰无误，逻辑连贯通顺，避免词汇的歧义，受众明确，以确保模型能够理解任务要求。

阶段 4 (结果评价): 语言准确性评价——评价生成的内容的语法、用词和句子结构的准确性，以及统计拼写和标点的使用不当情况；内容质量评价——考察内容的质量，即是否与主题相关，所传达的信息是否准确和文本创新性。语境适应性——评价生成内容的文风和语境表达，以及对自然情感的理解能力。

综上所述，生成式 AI 在数据分析与挖掘课程资源构建中的理论基础坚实且富有应用价值。通过深入挖掘生成式 AI 与教育资源构建的结合点，分析其在教育领域的应用价值以及与数据分析挖掘课程的契合度，我们可以为数据分析与挖掘课程资源构建提供新的思路和方法，推动教育教学的创新和变革。

4. 生成式 AI 辅助“数据分析与挖掘”课程资源构建的实践

本文分别针对教案生成选择行业内公认的工具文心一言、PPT 生成选择 Gamma、代码生成选择 Github Copilot 进行实践。

4.1. 教案生成——以文心一言为例

在提示词里输入：帮我写一个数据分析与挖掘课程的“数据预处理”的教案。运行文心一言，可得如图 3 和图 4 所示结果(限于篇幅，仅展示主体部分内容)。



Figure 3. The teaching plan generated by Wenxin Yiyang (Part 1)

图 3. 文心一言生成教案(部分 1)

从结果上我们可以看出：语言准确性方面，教案内容的语法、用词和结构的准确性都较高；拼写和标点没有使用不当的情况；内容质量较好，与主题完全吻合；语境表达与主题完全一致。所以，我们可以利用生成式 AI 较好地对外纲进行生成，然后对不够详细的地方进行修正，极大地方便了教学人员的备课工作。

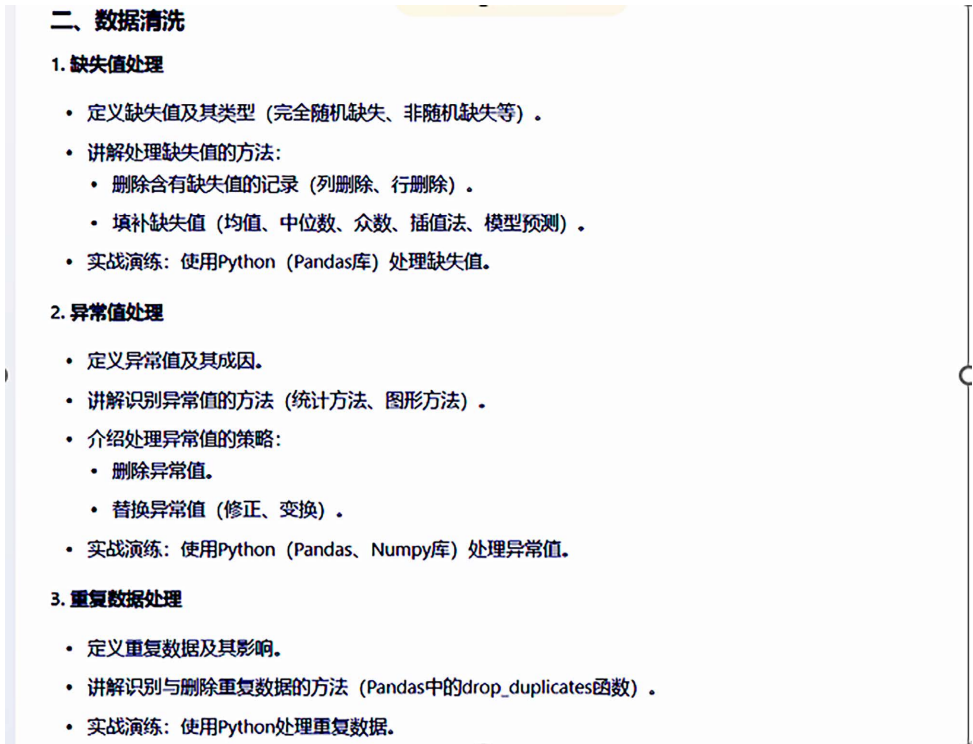


Figure 4. The teaching plan generated by Wenxin Yiyan (Part 2)
图 4. 文心一言生成教案(部分 2)

4.2. PPT 生成——以 Gamma 为例

在大纲输入框中输入如图 5 所属的 PPT 大纲，内容以第二章的“数据分析简介”为例。



Figure 5. Gamma PPT outline entry
图 5. Gamma PPT 大纲输入

选择“生成”，确定主题，然后选择“详细”进行生成，可得图6和图7所示的结果(限于篇幅，仅展示部分图片)。

从结果上我们可以看出：语言准确性方面，PPT 内容的语法、用词、结构、拼写和标点没有出现不当的情况；内容质量好，数据来源可靠，与主题较吻合；语境表达与主题一致；PPT 图文并茂，内容丰富且详实。所以，利用生成式 AI 可极大地提高教师备课效率。



Figure 6. The generated PPT by Gamma (Part 1)

图 6. Gamma PPT 生成结果(部分 1)



Figure 7. The generated PPT by Gamma (Part 2)

图 7. Gamma PPT 生成结果(部分 2)

4.3. 代码生成——以 Github Copilot 为例

以 Github Copilot 工具为例。Github Copilot 是一款快速且智能化的 AI 编程助手，是微软、OpenAI 和 GitHub 联合训练得到的代码编程大模型。通过使用 Github Copilot，我们可以直接生成想要的程序片段，

甚至是完整代码。首先通过在 Pycharm 里安装 Github Copilot 插件，并输入如图 8 所示的注释“#采用线性回归分析 data 数据，并计算准确率”。

```
1  -*- coding: utf-8 -*-
2  #逻辑回归 自动建模
3  import pandas as pd
4
5  #参数初始化
6  filename = './data/bankloan.xls'
7  data = pd.read_excel(filename)
8  x = data.iloc[:,8]
9  y = data.iloc[:,8]
10
11  #采用线性回归分析data数据，并计算准确率
```

Figure 8. Pycharm codes and comments entry

图 8. Pycharm 代码及注释输入

然后单击 Pycharm 的 Github Copilot 插件按钮，可得图 9 所示的生成结果。

```
Accept solution

print(u'通过随机逻辑回归模型筛选特征结束。')
print(u'有效特征为: %s' % ','.join(data.columns[
    rlr.get_support()]])
x = data[data.columns[rlr.get_support()]]
lr = LR()
lr.fit(x,y)
print(u'逻辑回归模型训练结束。')
print(u'模型的平均正确率为: %s' % lr.score(x,y))
```

Figure 9. Generated codes by Github Copilot

图 9. Github Copilot 代码生成结果

从结果上我们可以看出：生成代码基本满足我们的要求，与提示词(注释)的要求基本一致，只需少量修改即可满足我们的要求。代码生成的过程完美地匹配了上下文环境，极大地减少了代码书写量，提高了课程资源构建中代码书写的效率。

5. 未来展望

随着科技的飞速发展，生成式 AI 作为人工智能领域的重要分支，其在教育领域的应用逐渐凸显出巨大的潜力。特别是在数据分析与挖掘课程资源的构建中，生成式 AI 不仅提升了教学效率，还为教师提供了更加个性化和高效的备课体验。本文将对生成式 AI 在数据分析与挖掘课程资源构建中的应用进行深入

研究,并探讨其未来发展趋势以及课程资源的持续优化路径。

在未来发展中,生成式 AI 在数据分析与挖掘课程资源构建中的应用将更加广泛和深入。首先,随着技术的不断进步,生成式 AI 的处理速度和准确性将得到进一步提升,为课程资源的实时更新和优化提供有力支持。其次,生成式 AI 将与虚拟现实、增强现实等先进技术相结合,打造出更加沉浸式的学习环境,让学生在模拟实践中掌握数据分析与挖掘的核心技能。此外,生成式 AI 还将推动教育资源的共享和普及,打破地域和时间的限制,让更多人享受到优质的教育资源。

6. 结语

经过深入研究与分析,本文对于生成式 AI 在数据分析与挖掘课程资源构建中的应用进行了全面的探讨。在研究过程中,我们充分考察了生成式 AI 技术的特点及其在数据分析与挖掘领域的适用性,并结合实际课程资源构建的需求,进行了详实的理论分析和实证研究。在挖掘课程资源方面,生成式 AI 发挥出了显著的优势。传统的课程资源挖掘方法往往依赖于人工筛选和整理,效率低下且难以保证资源的质量。而生成式 AI 则能够通过智能算法,自动识别和挖掘出高质量的课程资源,大大提高了资源挖掘的效率和准确性。本研究的研究成果不仅丰富了相关领域的研究内容,还为后续的研究和实践提供了有益的参考和借鉴。展望未来,我们期待生成式 AI 能够在数据分析与挖掘课程资源构建中发挥更大的作用,为教育信息化的发展注入新的活力。

基金项目

重庆科技大学本科教育教学改革研究项目“新工科模式下‘智能科学与技术’新兴专业课程教学资源体系构建与探索”(202158);重庆科技大学本科教育教学改革研究项目“以‘课程育人’助推学生网络安全综合能力全面提升”(202463);重庆科技大学本科教育教学改革研究项目“基于 openGauss 的数据库原理及应用课程实验教学模式探索”(202464)。

参考文献

- [1] 孙道华,林丽芹,詹国武,等.生成式 AI 热潮给高等教育带来的挑战和思考[J].化工高等教育,2024,41(4):2-15.
- [2] 邓艺.用生成式 AI 设计“在线学习”教案——基于大模型的备课尝试[J].中国信息技术教育,2024,17(1):92-94.
- [3] 韦永圣,左佐堃川,陆飞宇.生成式 AI 技术在医学院校教学资源开发中的应用探析[J].电脑知识与技术,2024,20(27):31-33.
- [4] 张华锋.生成式 AI 辅助“数据结构”教学研究——以 CodeGeeX 为例[J].互联网周刊,2024,25(6):73-75.
- [5] 晏苏红,谢于晨,李卫勇.生成式人工智能驱动下的《信号与系统》教学改革研究[J].教育进展,2024,14(9):224-230.
- [6] 王宇轩,徐文浩,于浩淼,等.生成式 AI 为 C 语言编程教学带来的挑战和机遇[J].计算机教育,2024,8(1):133-141.
- [7] 李仲根,杨航,王青峰.生成式 AI 普及背景下环境科学与工程专业教育面临的机遇与挑战[J].教育观察,2024,13(10):15-19.
- [8] 舒莲卿,杜辉.ChatGPT 在课程教学中的应用初探——以 Python 程序设计为例[J].中国现代教育装备,2024(1):9-12.
- [9] 孙旭,钟秋菊,张文涛.生成式 AI 时代大学生智能学习助手:框架,挑战与应对[J].终身教育研究,2024,35(4):29-36.
- [10] 甄宏莉.人工智能背景下智慧课堂教学模式设计的行动研究[D]:[硕士学位论文].银川:宁夏大学,2020.
- [11] 戴韵.生成式人工智能技术赋能教师课堂教学质量评估[J].教育进展,2024,14(6):1264-1271.
- [12] 张龙.ChatGPT 生成式 AI 技术在信息教学中的应用[J].电子技术,2024,53(3):386-387.