



# 生成式人工智能辅助学科情报服务途径探析

——以利用 ChatGPT 生成学科领域论文分析报告为例

□江珊 常定姁 张开阳 龚芙蓉 张宁 盛芳\*

**摘要** 生成式人工智能的出现为图书馆的学科情报服务带来了新的机遇和挑战。文章基于工作实例,以 ChatGPT-4 为具体工具,探讨了利用生成式人工智能辅助制作学科领域论文分析报告的途径与方法。研究涵盖了从学科领域的确定到报告内容的生成,论述了在检索式生成、数据准备、数据分析与可视化、文献内容挖掘等环节中有效利用该工具的策略。研究结果表明,ChatGPT 在学科领域分析报告的多个阶段均能发挥重要作用,其灵活性、智能化与高效性显著提升了文献分析的深度和质量,为学科情报服务的优化提供了实证依据。

**关键词** ChatGPT 生成式人工智能 学科情报服务 论文分析

**分类号** G252

**DOI** 10.16603/j.issn1002-1027.2025.01.011

**引用本文格式** 江珊,常定姁,张开阳,等.生成式人工智能辅助学科情报服务途径探析——以利用 ChatGPT 生成学科领域论文分析报告为例[J].大学图书馆学报,2025,43(1):93-102.

## 1 引言

在数字化与信息爆炸的时代,图书馆如何在复杂情境与多元化技术手段中准确定位自身角色,成为亟待解决的问题。最有效的解决途径或许在于回归图书馆的本质职能,即守护知识并促进权威文献信息的有效利用。对于高校图书馆而言,这一职能尤为重要,正是基于此,许多高校图书馆推出学科情报服务<sup>[1]</sup>。学科情报服务的核心目标是促进权威来源文献信息的有效利用,通过主题检索和文献分析筛选出具备统计分析可靠性与有效性的文献样本数据集,据此客观揭示学科领域的研究热点,推送相关重要文献,整合交叉学科领域的研究成果,为研究人员、学生以及决策者提供深度的知识洞察。这一服务模式已经成为高校图书馆提升核心竞争力、增强用户满意度的重要举措之一。

然而,在开展学科领域文献分析的过程中,学科馆员面临两大挑战:一是如何尽可能获取准确而全面的文献数据集,二是如何突破现有分析工具的主题识别或挖掘的局限性。前者为正确分析提供基础,后者决定了能够揭示的知识深度。众所周知,文献分析对数据的准确性和全面性要求极高,但在实

际操作中,文献检索难以实现绝对的查全和查准。学科馆员只能尽可能减少主观偏差和知识局限,这无疑增加了分析结果的不确定性。此外,尽管许多数据库平台和分析工具已引入自动聚类 and 引文主题分类功能,但这些工具的分类体系、聚类算法以及可视化功能在理解和揭示复杂交叉学科知识方面仍存在一定局限性。近年来,生成式人工智能(Generative AI)在知识处理与文本生成领域展现出显著优势,能否利用这一技术克服文献分析中的上述挑战,成为一个值得关注和研究的问题。

## 2 生成式人工智能在文献分析中的研究现状

2022 年 11 月,OpenAI 公司推出了基于 GPT-3.5 架构的人工智能对话生成工具 ChatGPT,该工具一经问世便引发了广泛关注<sup>[2]</sup>,迅速使“生成式人工智能”这一概念得到广泛传播。以 ChatGPT 为代表的生成式人工智能能够自动生成文本、图像和其他形式的内容,显著减少人工整理信息的时间和成本,对知识的组织与呈现方式产生了深远的影响<sup>[3-4]</sup>。在科学研究领域,已有研究证明,使用生成式人工智能工具能够显著加速科学研究的进程。这些工具在生物医学、

\* 通讯作者:盛芳,邮箱:sf@lib.whu.edu.cn。



化学、物理学、地球科学、环境科学、天文学、地质学等领域均取得了令人瞩目的研究成果<sup>[5-11]</sup>。

在文献分析领域,ChatGPT 也展现了显著优势,能够用于自动生成文献综述、提取关键信息,并在一定程度上预测研究趋势。阿尔沙米(Alshami)和马辛德(Masinde)等学者利用 ChatGPT 自动化或简化系统性综述的生成步骤,有效缓解了传统系统性综述存在的主观性问题,可以帮助研究人员做出更全面、更可靠的文献综述<sup>[12-13]</sup>;德·温特(de Winter)等探讨了 ChatGPT 在科学计量学中的潜力,评估了 GPT-4 在预测引用次数、Mendeley 读者数量和社交媒体参与度方面的能力<sup>[14]</sup>;郑(Zheng)等利用提示工程引导 ChatGPT 进行金属有机框架的文本挖掘和合成预测,展示了 ChatGPT 在化学文献分析中的应用潜力<sup>[15]</sup>。

随着 ChatGPT 应用的不断扩展和效果的持续显现,关于生成式人工智能赋能图书馆各类服务的研究逐渐增多。目前,生成式人工智能在智慧图书馆框架构建<sup>[16-17]</sup>、提升服务效率和质量、提供个性化服务<sup>[18-19]</sup>等方面的显著优势已基本成为共识,未来的研究将进一步探讨其在具体服务场景中的赋能作用。在数据、知识和情报服务场景中,现有研究认为,图书馆应重视数据价值,并通过促进生态融合来适应新的信息环境<sup>[20]</sup>。生成式人工智能在科研知识生产中显示出积极作用,尤其是在知识传播和学术资源共享方面<sup>[21]</sup>。基于 GPT 技术的生成式人工智能工具在情感分析、数据整理、知识推理等方面展现出巨大潜力,能有效融入图书馆知识情报服务的各个环节,从战略决策、战术筹备到流程设计,深刻改变学术信息服务、决策情报服务和社会信息服务等多个层面<sup>[22-23]</sup>。此外,在应用研究方面,已有学

者通过构建文献数据集,与 ChatGPT 合作开展医学信息查找等研究,展示了生成式人工智能在文献分析和资源利用中的潜力<sup>[24]</sup>。上述研究为生成式人工智能赋能图书馆情报服务提供了理论基础和实践场景。

然而,尽管已有上述探索,生成式人工智能技术在图书馆文献情报服务中的应用仍然主要处于理论研究阶段,现有的实践场景研究在文献数据集规模、数据获取和分析利用等方面尚未能与学科领域文献分析的实际需求实现对接,特别是在如何利用生成式人工智能工具开展学科领域文献分析的全路径实践方面。鉴于此,本文将基于实际工作案例,探讨利用 ChatGPT 制作学科文献分析报告的途径与方法,重点分析其在解决文献分析关键难点时所发挥的作用,研究如何在学科情报服务的具体场景中有效整合生成式人工智能技术,以提升图书馆的知识服务能力和学科支持水平。

### 3 学科领域文献分析的内容及流程

学科领域文献分析通常基于宏观视角,揭示某领域的整体研究状况、发展趋势、研究热点和关键问题,内容通常包括文献数量与趋势分析、研究热点与主题分析、高影响力文献与作者分析、国际合作与地域分布分析、研究前沿与未来趋势预测、学术期刊与会议分析等。这些分析的结果通常以图表等可视化形式结合文字进行呈现。

图 1 展示了学科领域文献分析的基本工作流程,涵盖了研究领域确定、检索式构建、数据处理、报告形成六个阶段。

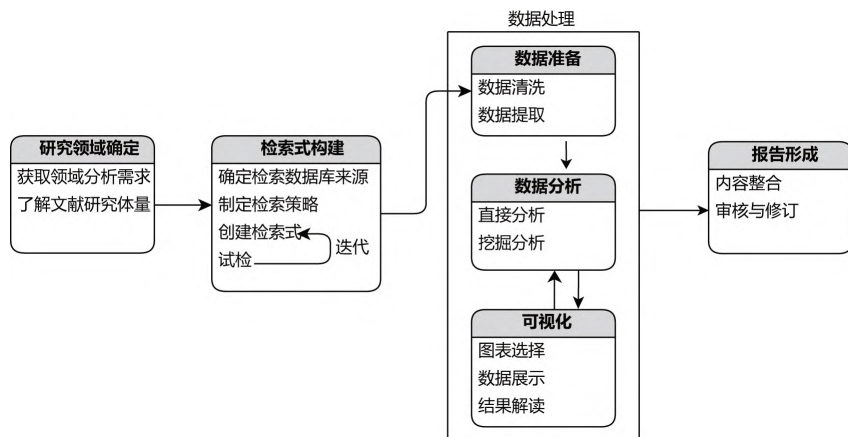


图 1 学科领域文献分析工作流程



首先,根据实际需求确定研究领域,并评估该领域的文献研究体量,以判断开展文献分析的可行性。在此基础上,通过构建检索式从相关数据库中获取文献数据集。检索式的构建需结合学科背景知识与信息检索知识,制定出具有针对性的检索策略,并在此基础上扩展检索词,针对具体数据库迭代试检,从而持续提升检索结果的全面性与准确性。

获取原始文献数据后,进入数据准备阶段。此阶段包括数据清洗与数据提取。数据清洗通常包括处理缺失值、去除重复项、标准化数据格式等;数据提取则指从源数据中提取特定的专项数据,以形成可供后续分析使用的专项数据表。例如,进行文献数量与趋势分析时,需要对原始数据中的“发表年”字段进行提取和同值聚类,形成文献年代分布表。

数据准备完成后,进入数据分析阶段。根据数据处理方式的不同,数据分析可分为直接分析与挖掘分析两类。直接分析主要针对显性数据,这些数据在文献题录中已有明确的对应字段,如发表年、发表期刊、引用次数等,可以直接从题录数据中统计获得。对应的分析内容包括文献数量与趋势分析、高影响力文献与作者分析、国际合作与地域分布分析、学术期刊与会议分析等;而挖掘分析则需对题录数据中的文本进行深入挖掘,提取出有价值的信息。这里的文本主要指题名、关键词和摘要,这三项字段在文献检索意义上构成了文章的主题,因此,挖掘分析主要对应主题相关内容,包括研究热点与主题分析、研究前沿与未来趋势预测等。

无论采用何种分析方式,结果的呈现都需选用合适的方法,以确保复杂数据能够被有效解读。相同的数据可以通过多种不同的揭示角度和呈现方式展现出不同的效果,因此,选择合适的可视化方式(如趋势图、网络图、热力图等)至关重要。最后,将各类分析结果整合成逻辑清晰的文档,并经相关领域专家审核与修订,最终形成一份基于可信数据的系统性的领域文献分析报告,为相关领域的科研与决策提供参考。

#### 4 利用 ChatGPT 辅助学科领域文献分析的实例与方法

本节将通过实例展示如何利用 ChatGPT 辅助实现前述工作流程,所采用的 ChatGPT 版本基于 GPT-4.0 模型。

##### 4.1 研究领域确定与检索式构建

###### 4.1.1 需求获取及可行性分析

高校图书馆的学科情报服务旨在推动教学科研、促进学科发展,因此校内各级单位的调研需求是学科领域分析的主要来源之一。本实例的学科领域为全球极地科研,来源为相关学院的学科发展需求。

传统的可行性分析通常包括:通过网络了解学科背景,在目标数据库中进行初检,并对检索结果进行粗略分析,以评估该主题的文献资源体量和学科交叉程度,从而确定其可行性及检索策略总纲。在这一阶段,对学科的揭示程度相对较低,因此传统的搜索引擎工具基本能够满足需要。但 ChatGPT 无疑能更高效地完成这一任务。同时,考虑到后续构建检索式的需求,本实例从初检阶段起开始引入 Chat GPT,以实现更流畅的交互和更连贯的分析。

###### 4.1.2 ChatGPT 辅助检索策略及检索式构建

初检结果的学科分布表明极地科研是典型的多学科交叉领域,不仅所涉学科众多,且存在许多新兴领域,因此,检索策略以“极地”和“科研”作为核心限定概念,通过 ChatGPT 从检索层面厘清这两个概念的内涵和外延。在实践中,本实例利用 GPT-4 的交互优化功能,通过背景知识获取、地域和科研限定、检索式搭建三步法,从一般性问题入手,逐步引导至具体的研究内容,从而激发 ChatGPT 的深层理解能力,生成更为精准的回应。

作为当前最先进的生成式人工智能之一,ChatGPT 经过大规模预训练,积累了丰富的知识概念,在常识性知识的全面性上已显著优于个体专家<sup>[25]</sup>。通过上述方案,ChatGPT 能够提供全面的视野,并随着问题的深入,更为精准地了解用户所指的科学领域,在检索词的扩展和检索式的构造上确保围绕“极地科研”核心范畴展开,最终获得基于明确检索式的完整的检索结果集。这种可解释性、可控性和与数据库的适配性相比于当前基于自然语言的 AI 检索更适合学科领域文献分析的场景。

##### 4.2 数据准备

ChatGPT 在数据准备阶段的辅助主要体现在两个方面。首先,GPT 具备直接的数据处理能力,能够对输入数据进行统一格式、修正拼写错误、转换大小写等文本标准化处理和数据提取,这一方式可称为“直接处理模式”。其次,GPT 具有强大的代码生成能力,能够根据具体需求生成具有各类相应功



能的代码<sup>[25]</sup>,从而实现数据的预处理,这一方式可称为“间接处理模式”。前者适用于少量数据,对于检索结果往往达数万条的文献分析而言,不具备足够的效率;后者则基本不受数据量限制,并且在处理时表现出灵活高效的特点。

例如,为分析不同国家/地区在南极、北极以及整个极地区域的科研情况,本实例从来源数据库中导出三个 CSV 格式的文献题录数据,然后将需求提交给

GPT-4。GPT-4 清晰整理出案例背景和个性化需求(见图 2 右上方),同时生成 Python 代码(见图 2 左侧)。图 2 右下方为该代码的输出结果片段。整个过程,包括编码与调试,耗时仅几分钟。

编程语言作为一种语言形式,具有明确的语法和语义规则,是 GPT 的核心优势之一。在数据准备阶段,尤其是数据提取环节引入 ChatGPT,能灵活满足交叉学科分析的复杂个性化需求。

```
Python.py
1 import pandas as pd
2
3 # 加载数据
4 south_pole = pd.read_excel('south_pole.xlsx')
5 north_pole = pd.read_excel('north_pole.xlsx')
6 south_north_pole = pd.read_excel('south_north_pole.xlsx')
7
8 # 提取所需的字段
9 south_pole = south_pole[['国家/地区']]
10 north_pole = north_pole[['国家/地区']]
11 south_north_pole = south_north_pole[['国家/地区']]
12
13 # 数据清洗: 处理缺失值
14 south_pole.dropna(subset=['国家/地区'], inplace=True)
15 north_pole.dropna(subset=['国家/地区'], inplace=True)
16 south_north_pole.dropna(subset=['国家/地区'], inplace=True)
17
18 # 分割国家/地区字段(假设以逗号分隔)
19 south_pole['国家/地区'] = south_pole['国家/地区'].str.split(',')
20 north_pole['国家/地区'] = north_pole['国家/地区'].str.split(',')
21 south_north_pole['国家/地区'] = south_north_pole['国家/地区'].str.split(',')
22
23 # 展平数据: 每行一个国家/地区
24 south_pole = south_pole.explode('国家/地区')
25 north_pole = north_pole.explode('国家/地区')
26 south_north_pole = south_north_pole.explode('国家/地区')
27
28 # 去除首尾空格并将国家/地区转换为标准格式
29 south_pole['国家/地区'] = south_pole['国家/地区'].str.strip().str.title()
30 north_pole['国家/地区'] = north_pole['国家/地区'].str.strip().str.title()
31 south_north_pole['国家/地区'] = south_north_pole['国家/地区'].str.strip().str.title()
```



#### 案例背景和个性化需求

##### 案例背景

用户拥有三个 Excel 表格文件,分别包含南极、北极和南北极地区的文献题录数据。每个表格的每一行代表一篇文献,包含题名、摘要、作者、作者关键词、发表年、被引频次、发文章刊、国家/地区等信息。用户希望对这些数据进行分析,以生成一个按国家/地区汇总的南北极、南极、北极文献数量分布,并按南北极文献数量的大小倒序排列。

##### 个性化需求

- 从三个 Excel 表格中提取数据。
- 按照国家/地区进行数据汇总。
- 统计南北极、南极和北极的文献数量。
- 生成一个按南北极文献数量倒序排列的结果表格。

国家/地区	南北极	南极	北极
美国	5073	2551	2449
加拿大	1978	423	1573
英格兰	1948	1271	701
中国	1730	872	860
德国	1522	866	649
挪威	1342	245	1122
俄罗斯	1194	246	967
澳大利亚	1041	870	193
法国	970	593	368
意大利	728	557	177
瑞典	590	159	439
日本	586	392	188
西班牙	561	401	162

图 2 GPT-4 辅助文献题录数据提取

#### 4.3 数据分析及可视化

在数据分析和可视化阶段,ChatGPT 不仅能显著提高效率,还能够突破用户的认知局限,提供更为深入和多维的分析视角。

##### 4.3.1 直接分析中的 ChatGPT 应用

高质量的数据分析是从大量数据中提取有用信息的再创造过程,需要借助合适的工具来完成。然而,传统的数据分析工具,如 Excel、Tableau、Power BI、Qlik Sense<sup>[26]</sup>等,虽然各具特色,但均依赖人工导向进行多角度尝试,以挖掘与实际数据最契合的分析视角。而以 GPT 为代表的生成式人工智能工具则通过引入“智能”改变了这一过程。具体而言,

分析者只需一次输入数据,ChatGPT 便能在后台利用各种算法进行全方位分析,自动提取最具揭示性的若干角度供用户选择。

本实例中,将表 1 所代表的年份数据导入 GPT-4,指示其同时生成多角度分析结果,并提供相应的图表和文字说明。几分钟内,GPT-4 输出了年度趋势分析、南北极比较分析、地区与科学发现关系分析、出版和时间的相关性分析四个分析结果,如图 3 所示。可以看出,GPT-4 能够有效理解数据的行列标题,为图表分析赋予明确涵义,展现出较高的分析价值。





表 1 某数据库极地科研发文章年代分布表(部分)

出版年	1991	1992	1993	1994	1995	1996	1997	1998
南北极	73	81	83	92	131	108	126	131
南极	46	44	54	44	73	62	61	74
北极	27	37	31	44	59	44	68	58
科学发现	27	30	24	20	53	36	42	38

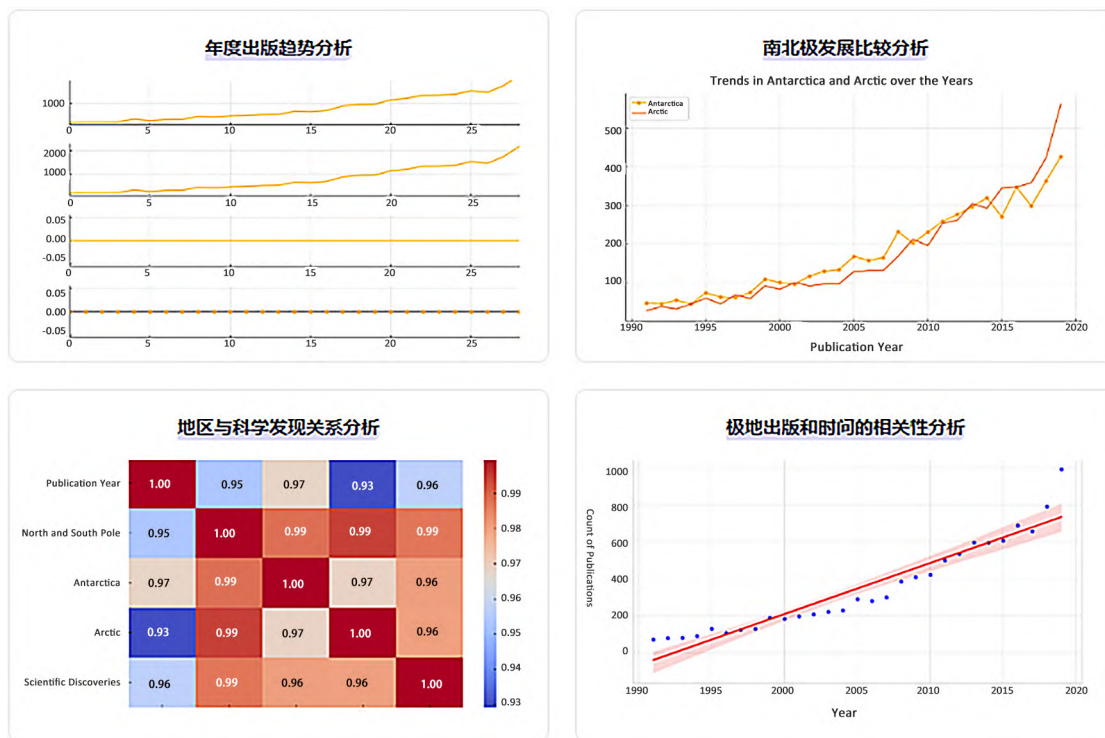


图 3 GPT-4 辅助文献年代分布分析

图 4 展示了将图 2 中的文献地域分布表输入 GPT-4 后的结果。GPT 根据地域分布数据的特点, 生成了南北极发文总量分析、南北极发文量对比分析、南北极发文量份额分析、南北极发文量趋势分析等几种输出。其中, 发文量趋势分析与图 3 中的年度出版趋势分析虽然本质上都是分析南北极发文趋势, 但所用图表和分析方法明显不同, 这表明 GPT-4 能够针对不同数据主题的特点进行选择, 提供更加多维的分析视角。

#### 4.3.2 挖掘分析中的 ChatGPT 应用

挖掘分析主要针对文献题录数据中的题名、关键词和摘要三个主题字段展开。本实例在实践中, 区分了关键词、题名、摘要的不同特点。

##### (1) ChatGPT 辅助关键词智能理解

由于不同作者的学术背景、研究领域及个人偏

好的差异, 关键词的拟定往往存在一定程度的随意性, 导致出现大量同义词和近义词。这一现象直接影响了数据分析的准确性, 而传统分析工具缺乏对同义词和近义词的“智能”识别能力。GPT-4 能够基于超越一般专家的知识储备, 实现同义词和近义词的识别与自动归一化处理, 使关键词频分析更具准确性。

如图 5 所示, 左表为利用 SciMAT<sup>[27]</sup> 提取的极地科研文献数据中的作者关键词及其词频统计片段, 共截取了 38 个原始作者关键词, 按词频降序排列。在此片段中, 多个近义词(例如“MOON”和“LUNAR”)并存。通过将左表及图 5 中间的提示文字输入 GPT-4, 生成了图中右表及下方的“归一化过程及其频次变化”文字说明。



图4 GPT-4 辅助文献国家/地区分布分析

结果表明,GPT-4 对该片段中的 6 组同义词或近义词进行了归一化处理,并按归一化后的词频重新降序排列。与左表中同义词和近义词散布的情况相比,右表表现出明显的聚类效果。文字说明部分则进一步增强了输出结果的可解释性,使用户能够在必要时进行干预。

#### (2) ChatGPT 辅助词云定制

在可视化方面,本实例通过 GPT-4 指导用户生成自定义词云图,来展示生成式人工智能工具带来的个性化体验。本阶段也利用了 ChatGPT 的代码生成能力。

将上节得到的归一化后的关键词表输入 GPT-4,并提出个性化要求,包括编程语言、开发环境,以及同时生成中英文两个词云图。GPT-4 同样在数秒内给出了详细的解决方案,包括如何在指定的运行环境中进行配置、安装中文字体、创建 Python 文件,以及具体的代码和执行步骤。图 6 展示了在 GPT-4 指导下,于 VSCode(指定的运行环境)中执行代码并获得的运行结果。这种完整的编码环境也为进一步的个性化需求提供了便捷的实现途径。

#### (3) 文献内容挖掘

传统文献分析由于受到高水平翻译和领域知识深度理解的限制,往往难以突破关键词层面,达到摘要及以上层面(如全文)的主题挖掘。以 GPT 为代表的大规模预训练语言模型极大地突破了这些障碍,尤其在跨语言和学科知识理解方面展现出显著优势。根据 OpenAI 发布的最新 GPT-4 技术报告,GPT-4 在 57 个学科专业的学术理解准确率达到 86.4%,常识性推理准确率达到 95.3%<sup>[28]</sup>。因此,ChatGPT-4 不仅能够消除语言障碍,还能快速提取文章的核心观点和关键结论,具备较高的准确率。

基于上述技术优势,本实例在“高影响力文献”栏目中利用 GPT-4 进行文献题名翻译,并根据摘要内容对涉及的极地问题进行简要解读,显著提升了文献推介的可读性。对于可能出现的理解偏差,通过在报告中标注解读来源和解读目的,帮助读者理性判断,审慎规避误读风险。此外,“高影响力文献”栏目的推荐量通常控制在 20 篇左右,这一输入量对于 GPT-4 而言处理效率较高,能够支持摘要层面乃至全文层面的分析。



图5 GPT-4 辅助关键词归一化词频统计

本实例中,“高影响力文献”栏目推荐了2020年以来极地科研领域被引频次排名前1%的论文,涵盖14个研究主题,共15篇文献。通过总体摘要分析,GPT-4 生成了一个研究概貌:“所列15篇文献中关于南极的有6篇,涉及南极冰盖、南极磷虾渔业和南极旅游、南极地区发现的酶、南极地区高能中微子探测等;关于北极的有9篇,涉及北极海洋、冻土、北极动物档案、北极微生物群落以及北极地区国家、居民与生态环境的交互影响”。此外,GPT-4 还对每篇文献进行简要概述,阐明该研究与南北极地域的具体关联。这种解读方式相较于直接翻译更为简明,且能揭示多学科文献在极地研究这一共同主题下的聚集原因,增强了报告的逻辑性和揭示深度。

## 5 ChatGPT 辅助文献分析的优势与途径总结

上节通过对极地科研领域文献分析的全流程探讨,详细阐述了 GPT-4 在学科情报分析各阶段的赋能作用。总体来说,ChatGPT 的优势功能与学科情

报分析场景的需求高度契合,整个过程主要利用了 ChatGPT 在文本理解生成和代码生成方面的两大核心功能。具体而言,可行性分析、检索式生成和高水平文献内容分析主要利用 ChatGPT 的文本理解生成能力;数据处理(包括数据准备、数据分析和可视化)则主要利用了 ChatGPT 的代码生成能力。上述功能的选择以文献分析各阶段的特点为依据,可以为生成式人工智能辅助学科情报分析提供一般性的思路参考。

在检索式生成方面,ChatGPT 帮助用户全面理解学科背景,明确检索策略,扩展领域术语和关键词,充当了学科知识辅助工具,极大缓解了本文开头提到的传统学科领域文献分析面临的两大挑战之一:如何获取尽可能准确而全面的文献数据集的问题。

在数据处理阶段,从数据准备到数据分析与可视化,ChatGPT 的赋能作用逐步增强。不仅能够显著提高工作效率,实现灵活定制,还能突破用户的经验局限,提升分析报告的知识揭示程度,在很大程度上





上缓解了传统分析面临的第二大挑战,即如何突破现有分析工具的主题识别或挖掘的局限性。对于这

一问题,GPT 不断增强的代码生成能力具有广阔的应用前景。



图6 GPT-4 辅助词云定制

在文本挖掘分析方面,针对关键词、题名、摘要等不同层级的分析,利用 ChatGPT 改良了传统关键词分析的揭示准确度,同时针对重要文献拓展了揭示深度,达到真正意义上的摘要及全文层面揭示。在这一环节,利用 GPT 强大的知识储备和机器学习能力进行交叉学科知识揭示无疑是未来值得关注的方向。

上述各环节共同构成了 ChatGPT 辅助学科领域文献分析的实现途径。该途径的优势在于充分考虑了分析需求和题录数据的特点,平衡了可解释性、计算资源和揭示深度等多种因素,实现了传统工具与生成式人工智能工具的优势互补,而非单纯依赖生成式人工智能进行一站式处理。当前,包括 Web of Science 在内的各大数据平台已纷纷引入 AI 功能,提供自然语言检索及分析服务,但这些方法在检索结果的可解释性、全面性和准确性方面目前尚无法满足学科领域文献分析的需求。针对海量文献题录数据,传统主题挖掘工具和基于 GPT 技术的生成式人工智能工具在主要性能指标上的对比见表 2

(由 GPT-4 生成),表中展示了两者在可解释性、准确性等方面的差异。从表 2 的结果可以看出,GPT 的优势在于其灵活性,特别是在处理多样化和非结构化数据方面。然而,在可解释性和计算资源需求上则逊色于传统工具。因此,本文提出的辅助方式是当前更为契合学科情报分析场景的生成式人工智能利用途径。

表2 GPT 与传统主题挖掘工具性能对比

特点/方法	传统工具	GPT
语义理解	有限	强
预处理要求	高	低
计算资源需求	低	高
灵活性	较低	高
可解释性	强	较弱
适用性	成熟的统计方法,适用于大规模数据	适用于多样化和非结构化数据,功能丰富

## 6 结语

虽然生成式人工智能在学科领域文献分析的过程中展现了灵活性、智能性和高效性,并在兼顾可





解释性的基础上拓展了分析的深度,但本方案仍存在一些不足之处。

首先,本方案对生成式人工智能工具的选择及使用要求较高。不同生成式人工智能工具在性能上存在较大差异,尤其是在知识理解和表达的准确性方面,这对学科情报分析至关重要。虽然当前市场上有许多生成式人工智能产品,但真正达到学科分析报告需求的却不多,这也是本文选择 ChatGPT 作为辅助工具的原因。在实际应用中,需审慎选择工具,同时始终保持谨慎态度,采用一系列验证措施确保分析过程在可解释和可控的范围内进行。

其次,本研究主要聚焦于整体路径的探索,对于一些复杂阶段的深入研究不足。例如,数据分析阶段涉及多种分析类型,在这一阶段,如何有效地将生成式人工智能工具与传统工具、甚至与其他生成式人工智能工具有机结合,以降低对硬件资源的依赖并提升分析的精度和深度,是研究的重点。

总之,生成式人工智能工具在学科情报分析领域已经展现出巨大的潜力。随着人工智能技术的不断进步,尤其是在可解释性、领域深度理解和数据更新等关键问题上的持续突破,未来这些技术和工具将在学科情报服务中发挥越来越重要的作用。在面对跨学科、动态发展的领域分析时,生成式人工智能将有助于扩展学科情报分析的边界,为学术界发现潜在创新点和新兴研究方向提供新的视角;在多模态数据集成方面,它将提升情报分析的时效性,通过综合文本、图像、实验数据等不同类型的信息,为复杂问题提供更加全面和深入的分析视角。未来,生成式人工智能工具将推动学科情报服务不断朝着更全面、更精准、更客观的方向发展。

## 参考文献

- 张蒂. 学科馆员——知识服务的践行者[J]. 图书馆工作与研究, 2012(2): 114—117.
- 澎湃新闻. ChatGPT 一周年|大模型新进化:智能体? 超级生态? 工具智能?[EB/OL]. [2024-09-19]. [https://www.thepaper.cn/newsDetail\\_forward\\_25470873](https://www.thepaper.cn/newsDetail_forward_25470873).
- 新华网. 这一年,人工智能“生成”精彩[EB/OL]. [2024-09-19]. <http://www.xinhuanet.com/tech/20231228/7917ea75586f40e3840d88df9b74f13/c.html>.
- 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743—750.
- Bort W, Baskin I I, Gimadiev T, et al. Discovery of novel chemical reactions by deep generative recurrent neural network[J]. Scientific Reports, 2021, 11(1): 15.
- Hsu C, Fannjiang C, Listgarten J. Generative models for protein structures and sequences[J]. Nature Biotechnology, 2024, 42(2): 196—199.
- Jin H, Kong F H, Li X Y, et al. Artificial intelligence in microplastic detection and pollution control[J/OL]. Environmental Research, 2024[2024-09-19]. <https://doi.org/10.1016/j.envres.2024.119812>.
- Liu Y, Yang Z W, Yu Z Y, et al. Generative artificial intelligence and its applications in materials science: current situation and future perspectives[J]. Journal of Materiomics, 2023, 9(4): 798—816.
- Taddeo M, Tsamados A, Cows J, et al. Artificial intelligence and the climate emergency: opportunities, challenges, and recommendations comment[J]. One Earth, 2021, 4(6): 776—779.
- Zuo R G, Carranza E J M. Machine learning-based mapping for mineral exploration[J]. Mathematical Geosciences, 2023, 55(7): 891—895.
- Farukhi S. Discovery of 69 new exoplanets using machine learning[R/OL]. [2024-09-19]. <https://newsroom.usra.edu/discovery-of-69-new-exoplanets-using-machine-learning/>.
- Alshami A, Elsayed M, Ali E, et al. Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions[J]. Systems, 2023, 11(7): 351.
- Masinde M. Enhancing systematic literature reviews using LDA and ChatGPT: case of framework for smart city planning[C]// Proceedings of the Africa Conference (IST), Electric Network, 2024.
- de Winter J. Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts[J/OL]. Scientometrics, 2024[2024-09-19]. <https://doi.org/10.1007/s11192-024-04939-y>.
- Zheng Z, Zhang O, Borgs C, et al. ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis[J]. Journal of the American Chemical Society, 2023, 145(32): 18048—18062.
- 赵杨, 张雪, 范圣悦. AIGC 驱动的智慧图书馆转型: 框架、路径与挑战[J]. 情报理论与实践, 2023, 46(7): 9—16.
- 郭利敏, 付雅明. 以大语言模型构建智慧图书馆: 框架和未来[J]. 图书馆杂志, 2023, 42(11): 22—30, 133.
- 陈莹. 风口或危机: 图书馆应用生成式人工智能的思考[J]. 图书馆学刊, 2024, 46(2): 91—94.
- 储节旺, 杜秀秀, 李佳轩. 人工智能生成内容对智慧图书馆服务的冲击及应用展望[J]. 情报理论与实践, 2023, 46(5): 6—13.
- 汪聪, 张春红, 高楠, 等. 面向人工智能生成内容时代的图书馆服务适应与创新[J]. 情报理论与实践, 2024, 47(7): 17—24.
- 储节旺, 杜秀秀. 生成式人工智能赋能科研知识生产研究述评[J]. 大学图书馆学报, 2024, 42(3): 108—117.
- 郭亚军, 马慧芳, 张鑫迪, 等. ChatGPT 赋能图书馆知识服务: 原理、场景与进路[J]. 图书馆建设, 2024(3): 60—68.
- 储节旺, 罗怡帆. 人工智能生成内容赋能图书馆知识服务的路径研究[J]. 情报理论与实践, 2024(8): 34—42.
- Lund B D, Khan D, Yuvaraj M. ChatGPT in medical libraries, possibilities and future directions: an integrative review[J]. Health Information and Libraries Journal, 2024, 41(1): 4—15.



- 25 Sébastien B V C, Ronen E, Johannes G, et al. Sparks of Artificial General Intelligence: early experiments with GPT-4 [J/OL]. arXiv, 2023 [2024-09-19]. <https://arxiv.org/abs/2303.12712>.
- 26 Qlik Technologies Inc. 可视化设计最佳实践 | Windows 版 Qlik Sense 帮助[EB/OL]. [2024-09-19]. <https://help.qlik.com/zh-CN/sense/May2024/Subsystems/Hub/Content/Sense-Hub/Visualizations/create-viz-best-practices.htm>.
- 27 Cobo M J, Lopez-Herrera A G, Herrera-Viedma E, et al. SciMAT: a new science mapping analysis software tool[J]. Journal of the American Society for Information Science and Technology, 2012, 63(8): 1609-1630.
- 28 OpenAI, Josh A, Steven A, et al. GPT-4 technical report [J/OL]. arXiv, 2023 [2024-09-19]. <https://arxiv.org/abs/2303.08774>.

作者贡献说明:

江珊:主导选题研究并作出整合性贡献,包括框架设计、数据分析和论文撰写

盛芳:策划研究选题,设计研究思路

常定炯、张开阳:数据调研、清洗、分析及研究过程的实施

龚芙蓉:文献检索与分析

张宁:文稿撰写、修订与补充实验验证

作者单位:武汉大学图书馆,湖北武汉,430072

收稿日期:2024年9月23日

修回日期:2025年1月8日

(责任编辑:支娟)

## Exploring the Approaches of Generative Artificial Intelligence in Assisting Subject Information Services

— A Case Study of Using ChatGPT to Generate Literature Analysis Reports in Academic Fields

JIANG Shan CHANG Dingxu ZHANG Kaiyang GONG Furong ZHANG Ning SHENG Fang

**Abstract:** The emergence of generative artificial intelligence has introduced unprecedented opportunities and challenges to subject information services in libraries. This study explores how to utilize generative artificial intelligence, particularly ChatGPT-4, to assist in the production of literature analysis reports in specific disciplines. The research mainly addresses the following questions: Can generative artificial intelligence solve the main difficulties of literature analysis? How can generative artificial intelligence improve the efficiency and depth of literature analysis? And how can libraries maximize the use of this technology to enhance the effectiveness of subject information services? This study aims to explore the specific application methods of generative artificial intelligence in subject literature analysis, covering the entire process from determining the subject area to generating the report content. The research focuses on analyzing the application of generative artificial intelligence in literature retrieval, data preparation, data analysis and visualization, as well as content mining, highlighting the advantages of this technology in overcoming the limitations of traditional tools. This study employed a case analysis method, systematically reviewing existing literature and combining practical cases to illustrate how generative artificial intelligence addresses challenges such as data accuracy, comprehensiveness of literature retrieval, and the limitations of existing analytical tools in subject information services. It points out that the two major advantages of ChatGPT—text understanding and generation, and code generation—are highly compatible with the needs of subject information analysis scenarios, and demonstrates through examples the functional selection and utilization of ChatGPT based on the needs at various stages of literature analysis, providing a general approach for using generative artificial intelligence to assist in subject information analysis. The research results indicate that ChatGPT can significantly improve the quality of literature analysis at multiple stages, providing more coherent and insightful analysis reports compared to traditional methods, especially excelling in interdisciplinary research. The study also finds that while generative artificial intelligence demonstrates high flexibility, intelligence, and efficiency in literature analysis, its application still faces challenges in selecting and verifying tools. To ensure the interpretability and reliability of the analysis results, the study recommends the integrated use of generative artificial intelligence with traditional tools, and even the combination of different AI models, to further enhance the accuracy and efficiency of the analysis. This research provides a theoretical and practical framework for understanding the role of generative artificial intelligence in subject information services.

**Keywords:** ChatGPT; Generative Artificial Intelligence; Subject Information Services; Paper Analysis