ChatGPT研究的网络可视化：一项基于术语与关键词共现网络分析

Deep Kumar Kirtania

Bankura Sammilani College馆员

deepkrlis@gmail.com

摘要：本文的主要目标是通过术语与关键词共现网络映射技术，识别ChatGPT的主要研究领域。为开展本研究，我们从Lens数据库中检索了共577篇出版物，用于网络可视化。研究结果显示，“chatgpt”的出现次数最多，其次是人工智能、大语言模型、GPT、研究等相关术语。本研究将有助于图书馆与信息科学领域以及计算机或信息技术领域的专业人士。

关键词：网络可视化、术语共现、关键词共现、人工智能、ChatGPT、文献计量学。

引言：文献计量网络可视化是指利用可视化工具和技术，基于引文、合著和共被引等文献计量数据，呈现并分析不同学术出版物之间关系的方法（Perianes- Rodriguez、Waltman & Van Eck，2016年；Najmi 等，2017年；Liao 等，2018年；Jiang、Ritchie & Benckendorff，2019年）。文献计量网络所采用的可视化技术包括图、地图和网络图，这些技术能够揭示研究领域、机构或个人学者的结构、演变和影响（Lee & Chen，2012年；Pradhan，2017年；Aparicio、Iturralde & Maseda，2019年）。这些可视化结果可以帮助研究人员识别其研究领域中的关键作者、重要出版物和新兴趋势，也可为希望评估学术研究影响力和作用的政策制定者及资助机构提供帮助（Ellegaard & Wallin，2015年；Reed 等，2021年）。文献计量网络可视化是理解学术出版物之间复杂关系的有力工具，能够为学术交流与合作的动态提供有价值的洞见（Zhang & Eichmann-Kalwara，2019年；Shasha 等，2020年）。术语与关键词共现是指某些术语或关键词在文本语料库中频繁共同出现的现象，这表明它们之间存在关系

或关联（Scott & Tribble，2006年；Hunston，2007年；Santoni & Pourabbas，

意义模式、深入了解不同概念与关键词之间关系的有力工具（Bondi，2010年；Cech，2017年；Radhakrishnan、Erbis、Isaacs & Kamarthi，2017年；Weerasekara 等．2022年）。这些网络可视化工具有助于分析 ChatGPT 关键词在学术出版物中的使用情况及其影响。

目标：本研究的主要目标是通过术语和关键词共现网络映射技术，确定 ChatGPT 的主要研究领域。

方法：The Lens（https://www.lens.org/）数据库检索了与 ChatGPT 相关的学术作品，并选取 577 项研究成果开展本研究。The Lens 或 Patent Lens 是一个在线专利与学术文献检索数据库及分析工具，可通过引文索引提供全球学术文献元数据。研究人员以 Excel 格式下载了这 577 篇出版物。随后，使用 VOSviewer（https://www.vosviewer.com/）软件对这些 Excel 文件进行网络分析，并从中提取关键词。最后，借助 VOSviewer 软件完成数据可视化分析，并利用 Lens 数据库生成主要研究领域的词云。

局限性：由于 ChatGPT 是一个非常新的主题，相关出版物数量不多。这是本研究的一项局限，因为此类研究的出版物越多，研究结果就越可靠。

结果的可视化

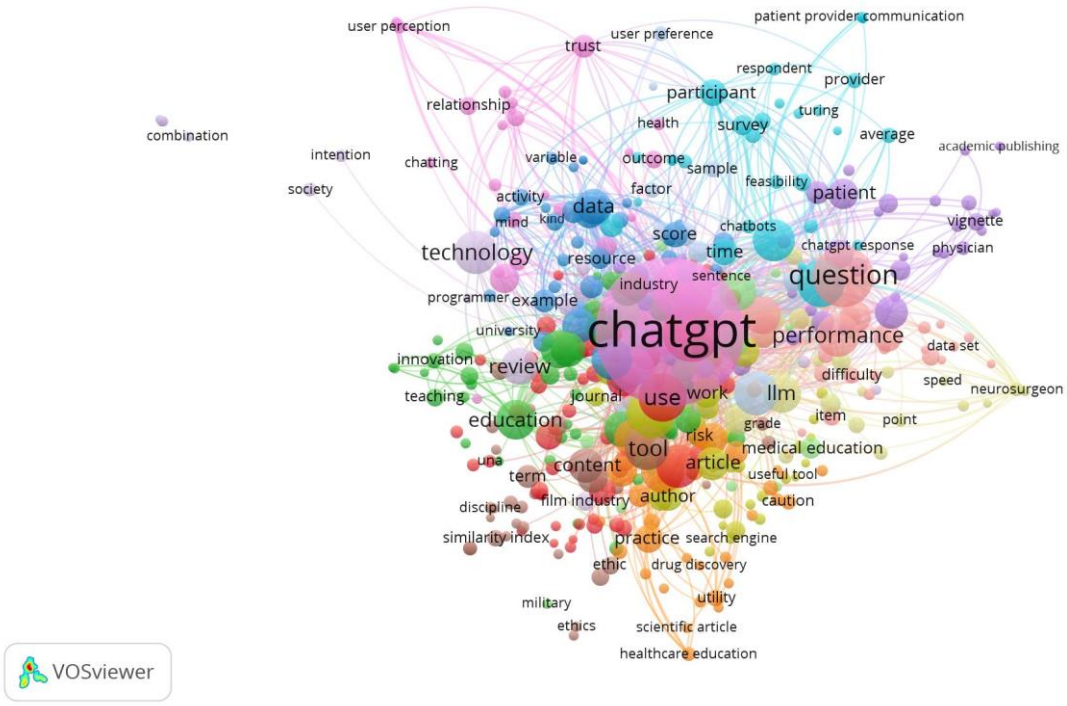


图 1：标题和摘要术语共现网络可视化

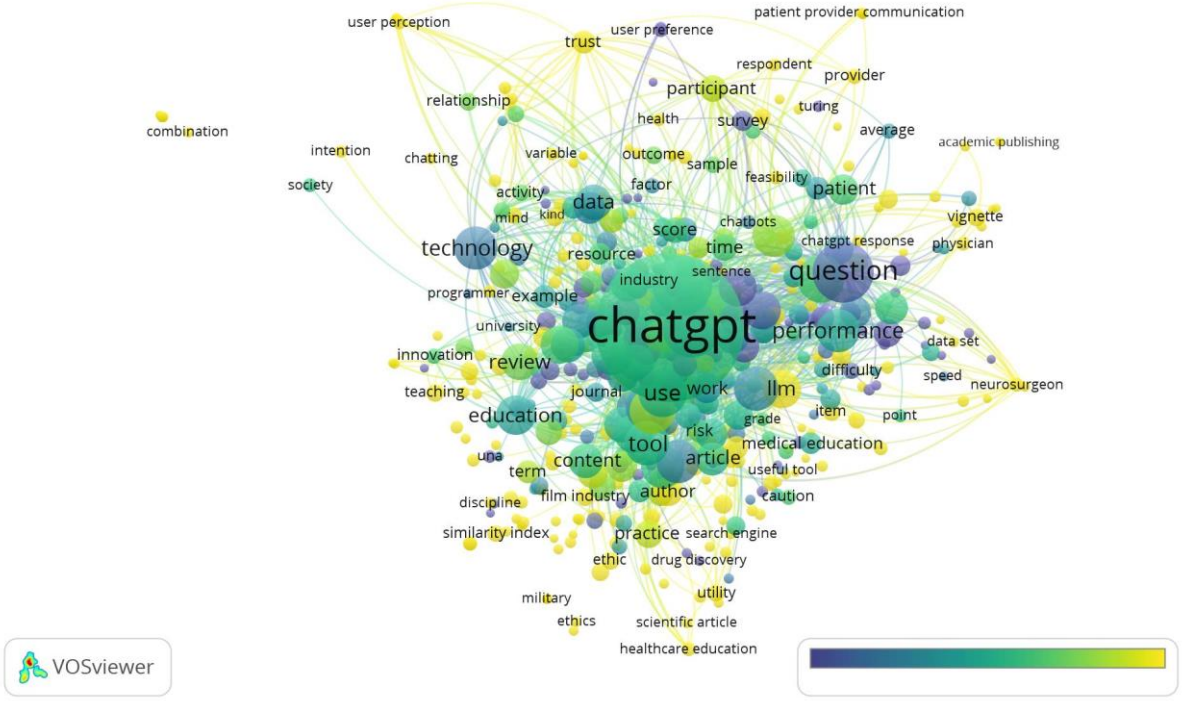


图 2：标题和摘要术语共现叠加可视化

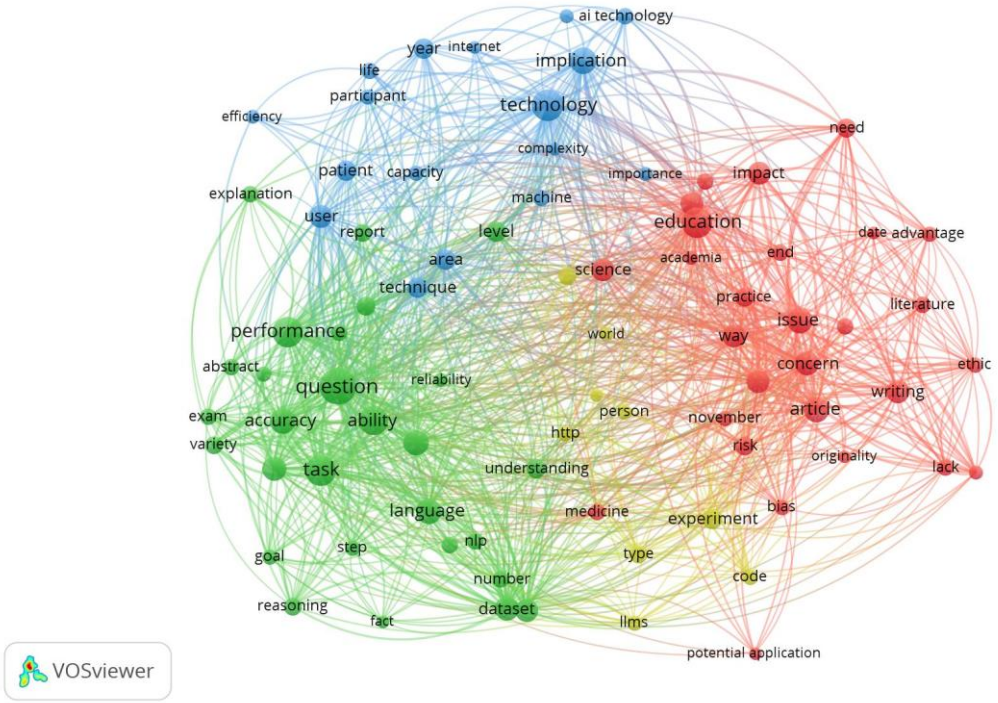


图 3：摘要术语共现网络可视化

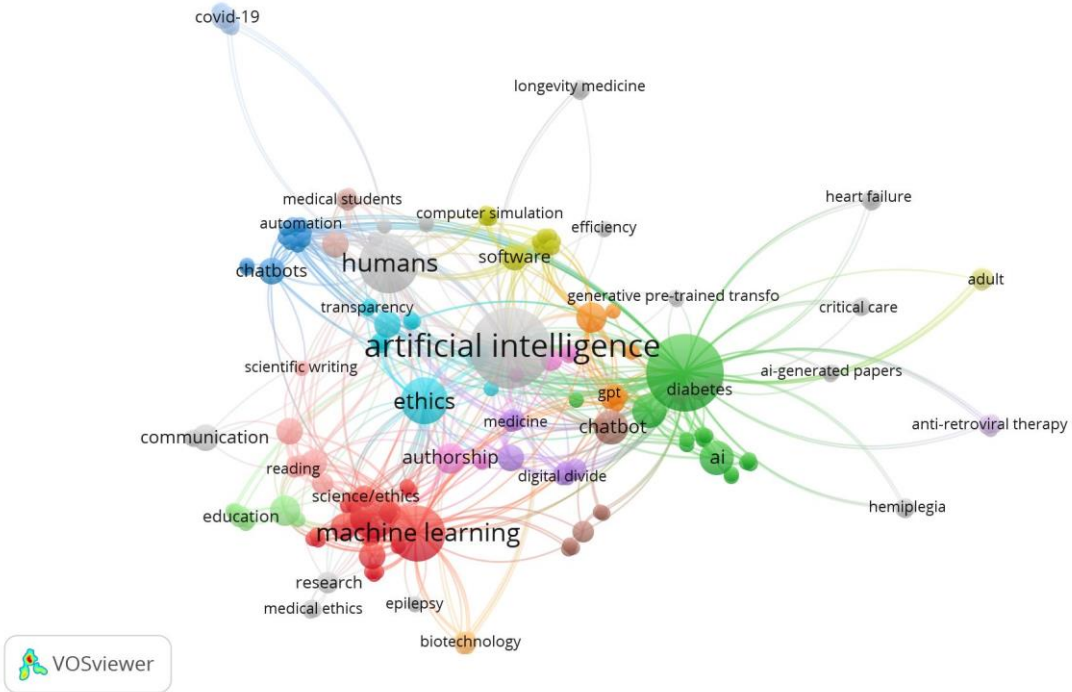


图 4：所有关键词的关键词共现网络可视化

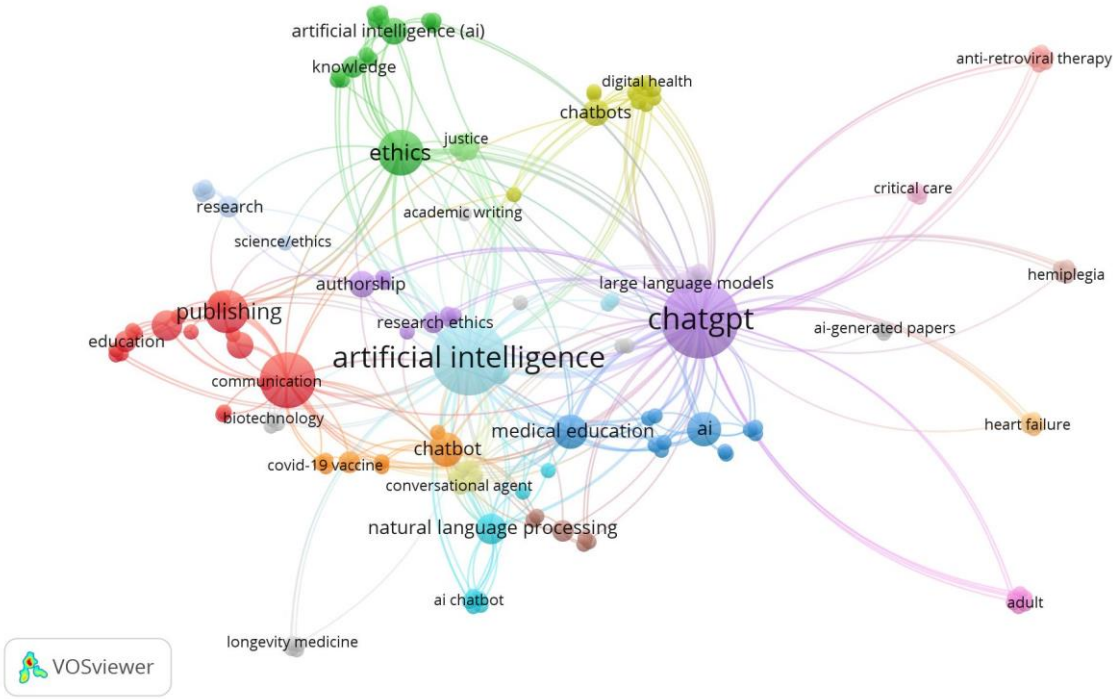


图 5：作者关键词的关键词共现网络可视化

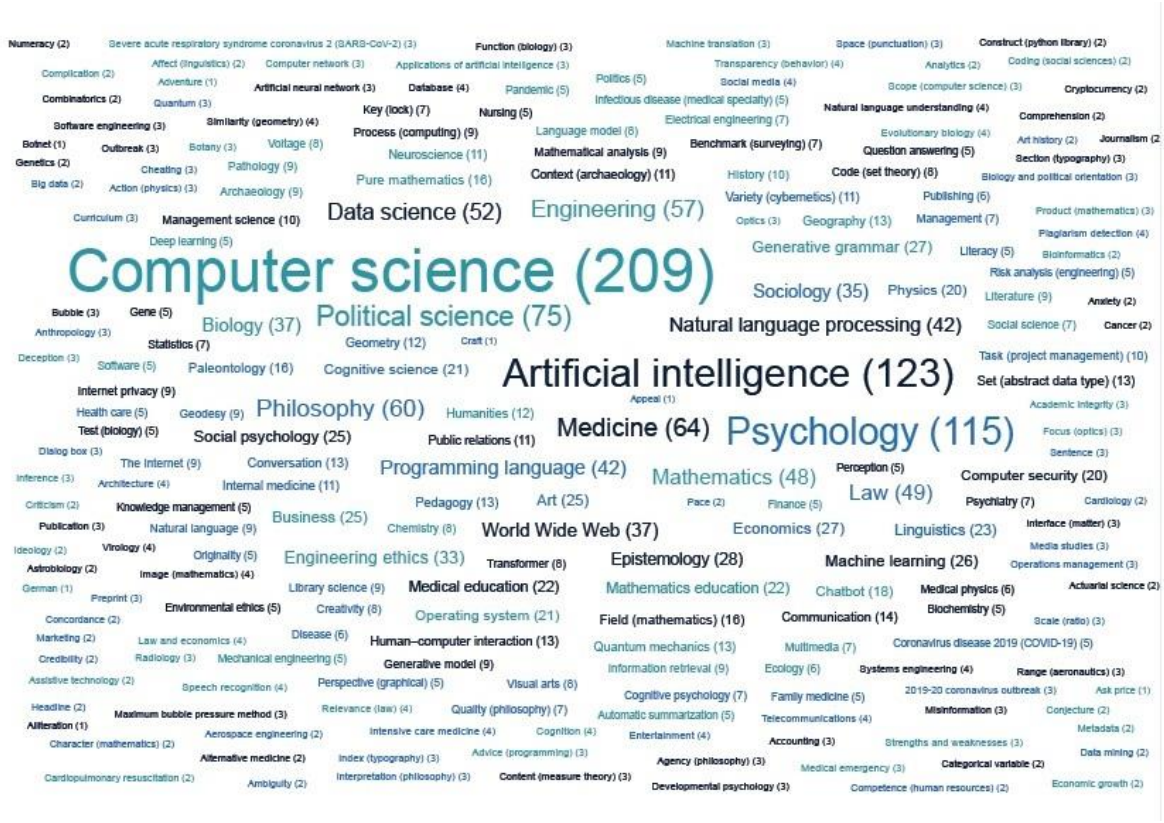


图 5：热门研究领域词云（**https://www.lens.org/**）

研究的主要发现

本研究的主要发现如下：

I. 从577篇出版物的标题和摘要字段中提取术语共现数据，共获得6908个术语。附录1列出了其中出现次数最多的50个术语，并提供了这些术语的出现次数、相关性得分和百分比。从表中可以看出，ChatGPT的排名最高，出现1276次，其次是问题（215次）、研究（208次）、模型（188次）等。这些术语之间的密切关系在网络可视化图像（图1、2和3）中一目了然。II. 所有关键词的关键词共现映射（图4）显示，人工智能排名第一，其次是ChatGPT、人类、机器学习、伦理学、出版等。III. 作者关键词的关键词共现映射（图5）显示，ChatGPT、人工智能和机器学习是使用最多的术语，其次是伦理学、出版、AI、聊天机器人、医学教育等。IV. ChatGPT的主要研究领域显示，大多数论文发表于计算机科学领域（209篇），其次分别是人工智能（123篇）和心理学（115篇）。此外，许多其他主题也已开始借助ChatGPT开展研究，Lens数据库提供的词云（图5）清楚地展示了这一点。从主题趋势来看，研究人员目前正在全球范围内研究ChatGPT在各种主题中的应用，未来也将围绕这一主题开展更多研究。

结论：网络可视化和共现网络分析已成为当前背景下非常重要的研究主题。借助这些方法，我们可以了解各种实体之间的关系，并轻松分析它们之间基于研究的关联。本文对所有已发表的 ChatGPT 相关出版物进行了可视化网络分析。本文表明，与 ChatGPT 相关的关键词使用最频繁，且这些词的出现频次和相关性得分最高。ChatGPT 是一个刚刚兴起的主题，仍有大量研究有待开展。可以预见，未来将会有更多关于 ChatGPT 的研究工作，

并对这些研究开展更多网络分析。也就是说，可以轻易断定，未来还可以围绕这一主题开展更多研究。

参考文献：

Aparicio, G., Iturralde, T., & Maseda, A. (2019年). 创业教育研究的概念结构与视角：文献计量综述
理与商[illegible]的视角. 文本中
词显著性, 1-20. Cech, T. (2017年). 探索社会信息学中的新兴主题：用于关键词共现分析的在线实时[illegible]
9月13日至15日，论文集，
[illegible]年). 学术成果的
量分析[illegible]adsen, D. Ø.
ChatG[illegible]
https://doi.org/10.20944/preprints202303.0479.v1. Hunston, S. (2007年). 重新审视语义韵. 《国际
[illegible]. 文献计量可视
旅游危[illegible], 22(16), 1925-1957. Kirtania, D. K. (2023年

https:/[illegible]T. T. (2012年)
识管理[illegible]领域的研究主题与趋势：从1993年到2010年. 《知识型系统》, 28, 47-58. Liao, H., Tang, M.,
[illegible]化. Sustainab
10(1), [illegible]-Infante, L., Adenso-Díaz, B., & García, S. (2019年). 对近期效率分
[illegible]H., Abbasi, A.
Travis[illegible]计量学》, 110,
865-[illegible]Valtman, L., & Van Eck, N. J. (2016年). 构建文献计量网络：全面论

Pradhan, P. (2017年). 文献计量学与科学计量学研究中使用的科学制图和可视化工具：概述。 Vol.23[illegible] Krishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017年). 基于新型关键词共现网络的方法，以促进科学[illegible]
Martin[illegible]t-Rolle, K., Dallmer, M., & Holden, J. (2021年). 研究影响评估：方法论框架。*Research Policy*，*50*(4), 104147. Santoni, D., & Poda[illegible]
Compu[illegible]t, M., & Tribble, C. (2006年). 文本模式：语言教育中的关键词与语料库分析（第22卷）. John Beniamins Publishing. Shasha. Z. Y., [illegible]与未来视[illegible]1年至2018年的文献计量综述.*Environmental Science and Pollution Research*，*27*，23514-23528. Weerasekara, S., Lu, Z., Ozek, B., Isaacs, J., & Kamar[illegible]的周期管理中的应用趋势：关键词共现网络综述与分析.*Sustainability*，*14*(19), 12233. Zhang, L., & Eichm[illegible]文献计量[illegible]
Communication，*7*(1), https://doi.org/10.7710/2162-3309.2266

附录1：标题和摘要中术语共现前50项

term	出现次数	相关性得分	百分比
ChatGPT	1276	0.0381	18.47
问题	215	0.2896	3.11
研究	208	0.1293	3.01
模型	188	0.1187	2.72
use	137	0.0585	1.98
响应	128	0.3439	1.85
人工智能	124	0.1263	1.80
大语言模型	116	0.0956	1.68

tool	116	0.1185	1.68
text	115	0.1253	1.66
gpt	111	0.4569	1.61
研究	110	0.2952	1.59
技术	110	0.3435	1.59
性能	108	0.2939	1.56
聊天机器人	103	0.4858	1.49
论文	100	0.1604	1.45
教育	97	0.6251	1.40
llm	88	0.1658	1.27
应用	87	0.1891	1.26
task	85	0.1249	1.23
信息	83	0.1857	1.20
系统	83	0.5392	1.20
评审	78	0.2389	1.13
学生	77	0.1696	1.11
data	75	0.2729	1.09
OpenAI	72	0.0973	1.04
挑战	71	0.1577	1.03
内容	69	0.5895	1.00
字段	67	0.25	0.97
role	67	0.2091	0.97
文章	66	0.2202	0.96
准确性	63	0.529	0.91
能力	62	0.1158	0.90
语言模型	61	0.0713	0.88
user	61	0.4332	0.88
分析	60	0.0793	0.87
局限性	60	0.122	0.87
答案	59	0.2812	0.85
影响	59	0.1055	0.85
方法	58	0.3713	0.84

能力	57	0.0927	0.83
患者	57	0.6724	0.83
问题	54	0.3323	0.78
知识	54	0.1308	0.78
科学	54	0.403	0.78
数据集	53	0.2038	0.77
背景	52	0.2209	0.75
作者	51	0.3966	0.74
case	51	0.4292	0.74
主题	51	0.1271	0.74