图书情报知识
Documentation,Information & Knowledge
ISSN 1003-2797,CN 42-1085/G2

《图书情报知识》网络首发论文

题目：ChatGPT 为代表的大模型对信息资源管理的影响
作者：陆伟，刘家伟，马永强，程齐凯
网络首发日期：2023-02-26
引用格式：陆伟，刘家伟，马永强，程齐凯．ChatGPT 为代表的大模型对信息资源管理的影响[J/OL]．图书情报知识.
https://kns.cnki.net/kcms/detail//42.1085.G2.20230224.1136.002.html

网络首发时间：2023-02-26 20:57:52
网络首发地址：https://kns.cnki.net/kcms/detail//42.1085.G2.20230224.1136.002.html

ChatGPT为代表的大模型对信息资源管理的影响

The Influence of Large Language Models Represented by ChatGPT on Information Resources Management

陆伟[1,2] 刘家伟[1,2] 马永强[1,2] 程齐凯[1,2]
LU Wei LIU Jiawei MA Yongqiang CHENG Qikai

（1.武汉大学信息管理学院，武汉，430072；2.武汉大学信息检索与知识挖掘研究所，武汉，430072）

摘要：OpenAI 发布的新一代对话型语言模型——ChatGPT，展现了自然流畅的对话能力。原本被认为不太可能的通用人工智能曙光重现。以 ChatGPT 为代表的大模型是数智时代的典型技术和应用创新。面对 ChatGPT 强大的信息加工、荟萃、整合和生成能力，信息资源管理学科机遇与挑战同在。ChatGPT 在信息资源管理支撑算法与技术、信息资源建设、信息组织与信息检索、信息治理、内容安全与评价、人机智能交互与协同等方面都具有深远的影响。数智时代，人工智能大模型飞速发展，我们有必要对此保持密切关注。依托以 ChatGPT 为代表的大模型，通过学科技术应用范式转换、理论方法创新、治理变革，可以进一步夯实信息资源支撑“四个面向”的基础。

关键词：ChatGPT；大模型；信息资源管理
中图分类号：G25

Abstract: OpenAI has released a new generation of conversational language model: ChatGPT, showing natural, fluid conversation capabilities, reviving the promise of artificial general intelligence that was previously thought impossible. Large language models represented by ChatGPT are typical technological and application innovations in the era of digital intelligence. ChatGPT's powerful ability of information processing, collecting, integrating and generating brings challenges and opportunities to information resource management discipline.It has had a profound impact on the six perspectives of information resources management including supporting algorithms and technologies, information resources construction, information organization and information retrieval, information governance, content security and evaluation, and human-computer intelligent interaction and collaboration. In the era of digital intelligence, with the rapid development of large-scale AI models, it is necessary for us to keep close attention to this and promote the corresponding transformation of subject technology application paradigm, innovation of theories and methods, and governance reform, so as to further consolidate the foundation of information resources to support the "Four Orientations".

Keywords: ChatGPT; Large language model; Information resources management

作为新一代专注于对话生成的语言模型[1]，ChatGPT能够根据用户的输入文本，利用自身强大的自然语言理解和生成能力产生自然流畅的回答，很多时候，还可以在多轮实时互动过程中给出问题的合理答案。一经推出，月活用户量迅速过亿[2]，更是引发了各界广泛热烈的讨论。除了具备根据上下文进行多轮对话问答的能力，ChatGPT还在信息抽取、文章撰写、代码生成、自动摘要、翻译等场景展现了出色的性能，使得原本被认为不太可能的通用人工智能重新显现了希望，也带来了人机智能交互与协同新突破。比尔盖茨甚至认为ChatGPT这类技术将变得和PC互联网一样重要[3]。下面我们将对ChatGPT的本质及其系列模型核心技术特征演进路径进行分析，并以此为导线探析大模型对信息资源管理学科研究与实践带来的影响。

ChatGPT本质上是一个基于生成式预训练语言模型GPT（Generative Pre-trained Transformer）[4]进一步开发的对话式生成模型。GPT与BERT[5]这两类预训练模型都采用了Transformer[6]作为底层结构，但BERT使用的是Transformer的Encoder，属于双向语言模型，而GPT则使用Decoder进行了单向语言模型的预训练。GPT系列模型主要包含GPT[4]，GPT-2[7]，GPT-3[8]，InstructGPT[9]，GPT-3.5和ChatGPT。为了便于比较和理

[基金项目] 本文系国家自然科学基金重点项目“数智赋能的科技信息资源与知识管理理论变革”（72234005）和国家自然科学基金面上项目“基于机器阅读理解的科学命题文本论证逻辑识别”（72174157）的研究成果之一。（This is an outcome of the key project "Data and Intelligence Empowered Theoretic Change of Scientific Information Resource and Knowledge Management Theory"(72234005) and the project "Argumentation Logic Recognition of Scientific Proposition Text based on Machine Reading Comprehension"(72174157) ,both supported by National Natural Science Foundation of China.）

[通讯作者] 陆伟（ORCID:0000-0002-0929-7416），博士，教授，研究方向：信息检索、AI治理、人机协同，Email: weilu@whu.edu.cn。（Correspondence should be addressed to LU Wei, Email:weilu@whu.edu.cn, ORCID: 0000-0002-0929-7416）

[作者简介] 刘家伟（ORCID:0000-0002-2774-1509），博士研究生，研究方向：信息检索、信息安全，Email: laujames2017@whu.edu.cn; 马永强（ORCID:0000-0002-4980-9834），博士研究生，研究方向：信息抽取、文档智能，Email: mayq97@qq.com; 程齐凯（ORCID:0000-0003-3904-8901），博士，副教授，研究方向：文本挖掘、信息检索，Email:chengqikai@whu.edu.cn。

解ChatGPT演进过程[10]，图1梳理了GPT系列模型在各个阶段的典型特征和关键改进。

OpenAI目前还未发布ChatGPT对应的论文。综合各方面的信息，ChatGPT是以GPT-3.5为底座，引入基于人类反馈的强化学习（Reinforcement Learning with Human Feedback,RLHF）[11]和高质量人机对话数据，通过大规模分布式集群训练得到。

ChatGPT成功的关键之一是使用了超大规模的预训练语料并拥有超千亿规模的模型参数。根据Chung等学者的研究[12]，模型参数规模在大于62亿的情况下，才能涌现出之前较小模型不具备的能力，而ChatGPT的参数据估计是1750亿。基于海量规模的语料训练，并应用所谓的上下文学习机制（In-Context Learning），ChatGPT可以适应广泛的下游任务，在低资源和零数据场景下有较好的语言理解和生成能力。

除此之外，ChatGPT另一个重要工作是引入了RLHF，利用人类的偏好作为奖励信号来微调模型，使得模型生成内容与人类常识、认知、需求、价值观保持一致，理解人类语言和完成人类指令，由此生成的回复符合人的选择偏好。同时，这也让ChatGPT与之前的大模型相比，其对话生成实现了从命令驱动到意图驱动的转换。

在此基础上，ChatGPT另外一个关键点是使用了高质量和多样化的数据，包括OpenAI搜集的历史对话数据、人工精细化标注的多轮对话数据和候选项比较排序数据。通过高质量多样化的数据微调、偏好奖励引导优化，让模型能够充分理解人类指令输入的意图，也为提升模型训练稳定性和性能上限提供了支持。

基于对ChatGPT本质及其系列模型核心技术特征演进路径的分析，我们从支撑算法与技术、信息资源建设、信息组织与信息检索、信息治理、内容安全与评价、人机智能交互与协同六个角度探析大模型对信息资源管理学科研究与实践带来的影响，形成以下观点：

（1）支撑算法与技术

现有各类抽取、识别、分类、生成等细分领域的大部分任务主要使用大量标注数据、训练或微调的范式，大模型技术的探索和应用相对较少，性能和效率还存在较大进步空间。未来需要优化大数据的使用，完善相应的数据质量和效用评价机制，在大模型底座的加持下，要探索更加高效的算法，用好“小”而“精”的高质量领域数据。此外，随着以ChatGPT为代表的大模型带来的能力突破，各类抽取、识别、分类、生成等领域细分的大部分任务不再需要花费大量资金和人员去进行大规模标注，而是要探索运用低资源、零样本提示学习思想，研发高效可迁移的算法技术，基于领域知识设计相应的指令，直接让大模型生成任务结果，让科研试错更高效，集中更多精力在机理机制的发现和分析上。

ChatGPT“黑盒”式的生成机制大大地削弱了其在关键场景的应用价值。例如，在辅助梳理相关研究的场景中，有研究发现[13]，ChatGPT生成的文献综述，文字

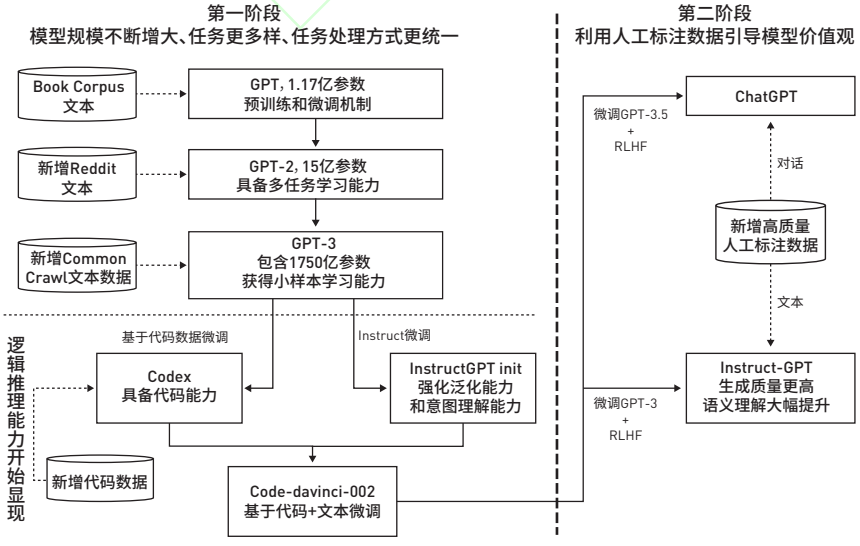

图1 ChatGPT各个阶段的典型特征和演进路径

Fig. 1 Typical Characteristics and Evolution Paths of ChatGPT at Each Stage

整体看起来流畅，但是其中包含的参考文献可能实际并不存在。人在总结相关工作时会有严谨的参考逻辑，而机器目前更多是依赖从学习过的语料中回忆，其生成过程和生成逻辑无法鉴别，未来需要探索相关算法，进一步提升ChatGPT类模型生成过程和内容的可解释性。

（2）信息资源建设

ChatGPT类大模型带来了颠覆性的多源多模态信息汇聚与生成能力，推动了信息资源建设和AI内容生成的技术升级。ChatGPT类大模型覆盖了海量的知识资源，其本身就是具备全域数据汇聚融合推理能力的新型信息源，未来需要探索针对这一新型数据源的建设组织与应用模式。ChatGPT类大模型还提供了强有力的数据关联、任务解决和内容生成能力，推动形成了未来人类生成内容和AI生成内容并存的新型信息环境。如何针对这一环境开展信息资源建设、构建“面向未来”的新理论、新方法体系[14-15]，将成为一个关键科学问题。ChatGPT类大模型在信息推理、数据整编、报告撰写、知识库构建等方面的能力也进一步推动了衍生信息资源的建设。

（3）信息组织与信息检索

ChatGPT的出现将推动现有的信息组织与信息服务模式转型。如何对ChatGPT生成的信息形成资源化利用，如何提供更加个性化的信息资源服务，从而进一步夯实信息资源支撑“四个面向”的基础，是需要面对的挑战。为此需要提出新的信息描述框架和组织模式，在AI生成内容快速增长的未来，构建面向多模态信息的增量式信息描述框架与组织模式，具备从互联网、领域数据库等复杂来源持续记录和描述的能力。通过语义层面对多源多模态信息进行关联，实现大规模高质量动态资源的有效利用。此外，由于ChatGPT等主流大模型不具备动态信息持续更新功能，如何与传统搜索引擎结合，从知识细粒度智能理解、可靠可信可解释检索的角度，提高信息组织与信息检索的效果，实现由“检索+推荐”模式到“感知+检索+推荐+生成”模式的转变[16]。

（4）信息治理

ChatGPT类大模型的规模化应用会带来人工生成信息的爆炸。面对海量、来源不清、真假难辨的信息，人类面临的信息过载和信息噪声问题将会更加严重。在大模型生成内容快速增长的未来，对此类信息的加工、组织、评价和鉴别面临极大的挑战。在知识产权方面，ChatGPT等AI模型生成的内容所有权归属目前没有明确的法律法规说明。近期的AI会议如ICML 2023[17]和ACL 2023[18]等，期刊如Nature[19]等，都及时更新和增加了关于AI协助写作内容的政策，要求不能使用由ChatGPT或任何其他人工智能工具生成的文本。未来可能需要进一步解决语料库版权问题，用户和平台版权归属问题，以及研发新的文字水印[20]技术来明确知识产权。与此同时，大模型中可能包含的价值观和思维偏见会被恶意利用，就宗教、民族、人权等问题进行信息污染[21]，因此需要推进数据采纳与算法公平性研究，同时完善相应的内容审核制度。考虑到AI生成内容带来信息伦理问题，未来还需要推进与哲学、法学和社会学等学科的跨学科研究，讨论明晰针对AI生成内容的伦理观，优化信息治理能力。

（5）内容安全与评价

尽管ChatGPT通过强化学习模型一定程度过滤了仇恨、种族歧视、涉恐涉暴涉政等有毒或敏感生成内容，但这一能力仍然存在不足，存在“越狱”的风险，可能会被不法使用[22]。例如，ChatGPT可能被用于伪造高可信钓鱼邮件、发起社交网络攻击、盗取隐私、传播错误信息和信息操纵[23]。未来，需探索如何帮助用户有效鉴别ChatGPT生成的“Deepfake”[24]式虚假信息，以及鉴别机器生成的情绪化、偏见等[25]不和谐内容，推进针对AI生成内容的多维度评价研究[26-27]，从信息内容自身质量、信息来源和信息获取途径等阶段评价信息安全性和可靠性，绿色、高效地保障内容安全。

（6）人机智能交互与协同

ChatGPT作为预训练通用语言模型，对情感、暗示等人因信息仍然无法有效处理，暂时缺乏与语音、视觉、触觉、脑电等信息的交互能力。未来还需要探索多场景、多模态输入输出，以及人因信息兼容的新型模型，推动更高水平的人机协同、人智协同。ChatGPT类大模型的进步还催生了新的用户行为模式，需分析用户感知和认知因素，探索人机共生环境下人机协同行为模式，从而进一步优化智能信息服务、用户隐私保护。需要强调的是，ChatGPT类大模型提供了打通人类智能和机器智能信道屏障的可能性，有助于实现人类智能和机器智能的深度融合，带来的潜在效益可能远超于其在问答和任务解决上的意义。

总之，以ChatGPT为代表的大模型是数智时代的典型技术和应用创新，其强大的信息加工、荟萃、整合

和生成能力极大地加快了信息空间中信息资源的流动和循环速率，对信息资源管理学科研究和实践带来了挑战和机遇。我们有必要对此保持密切关注，化解挑战，抓住机遇，促进相应的技术应用范式转换，推动信息资源管理学科理论方法创新和治理变革，更高效地数智赋能行业，以推动智慧图书馆、情报智能、智慧档案、语义出版和数字人文等领域的快速发展，提升服务质量和效率。

作者贡献说明

陆伟：提出研究思路，设计研究方案，论文修订与定稿；

刘家伟，马永强：设计研究方案，收集和分析资料，论文撰写与修改；

程齐凯：设计研究方案，论文修订与定稿。

参考文献

[1] ChatGPT: Optimizing Language Models for Dialogue [EB/OL]. [2023-02-09]. https://openai.com/blog/chatgpt/.

[2] ChatGPT Sets Record for Fastest-growing User Base - Analyst Note [EB/OL]. [2023-02-09]. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

[3] Bill Gates Calls ChatGPT 'Every bit as Important as the PC' or the Internet [EB/OL].[2022-02-23]. https://www.businessinsider.com/bill-gates-chatgpt-ai-artificial-intelligenct-as-important-pc-internet-2023-2.

[4] Improving Language Understanding by Generative Pre-Training [EB/OL]. [2023-02-22]. https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf.

[5] Devlin J,Chang M W,Lee K,et al. BERT:Pre-Training of Deep Bidirectional Transformers for Language Understanding [EB/OL]. [2023-02-22]. 2018:arXiv:1810.04805. https://arxiv.org/abs/1810.04805.

[6] Vaswani A,Shazeer N,Parmar N,et al. Attention is all You Need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach,California,USA. New York:ACM,2017:6000–6010.

[7] Radford A,Wu J,Child R,et al. Language Models are Unsupervised Multitask Learners[J]. OpenAI blog, 2019, 1(8):9.

[8] Brown T B,Mann B,Ryder N,et al. Language Models are Few-Shot Learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver,BC,Canada. New York:ACM,2020:1877–1901.

[9] Ouyang,Wu J,Jiang X,et al. Training Language Models to Follow Instructions with Human Feedback[EB/OL]. [2023-02-22]. https://arxiv.org/abs/2203.02155.

[10] How does Gpt Obtain its Ability? Tracing Emergent Abilities of Language Models to Their Sources [EB/OL]. [2023-02-22]. https://yaofu.notion.site/How-does-GPT-Obtain-its-Ability-Tracing-Emergent-Abilities-of-Language-Models-to-their-Sources-b9a57ac0fcf74f30a1ab9e3e36fa1dc1.

[11] Christiano P,Leike J ,Brown T B,et al. Deep Reinforcement Learning from Human Preferences[EB/OL]. [2023-02-22]. https://arxiv.org/abs/1706.03741.

[12] Chung H W, Hou L, Longpre S, et al. Scaling Instruction-Finetuned Language Models[EB/OL]. [2023-02-22]. https://arxiv.org/abs/2210.11416.

[13] Ma Y Q,Liu J W,Yi F,et al. AI Vs. Human—Differentiation Analysis of Scientific Content Generation[EB/OL]. [2023-02-22]. https://arxiv.org/abs/2301.10416.

[14] 马费成 . 凝聚共识，推动信息资源管理一级学科建设 [J/OL]. 信息资源管理学报 .[2023-02-22]. http://kns.cnki.net/kcms/detail/42.1812.G2.20221019.1758.002.html.(Ma Feicheng. Building Consensus and Promoting the First-Level Discipline Construction ofInformation Resource Management[J/OL]. Journal of Information Resources Management.[2023-02-22].http://kns.cnki.net/kcms/detail/42.1812.G2.20221019.1758.002.html.)

[15] 冯惠玲 . 以信息资源管理的名义再绘学科蓝图 [J]. 信息资源管理学报 ,2022,12(6):4-10.(Feng Huiling. Redrawing the Disciplinary Blueprint in the Name of Information Resources Management[J]. Journal of Information Resources Management,2022,12(6):4-10.)

[16] 陆伟，杨金庆 . 数智赋能的情报学学科发展趋势探析 [J]. 信息资源管理学报 ,2022,12(2):4-12.(Lu Wei,Yang Jinqing. Exploration on the Development Trend of Information Science in the Era of Data Intelligence Empowerment[J]. Journal of Information Resources Management,2022,12(2):4-12.)

[17] Clarification on Large Language Model Policy LLM [EB/OL]. [2023-02-09]. https://icml.cc/Conferences/2023/llm-policy.

[18] ACL 2023 Policy on AI Writing Assistance [EB/OL]. [2023-02-09]. https://2023.aclweb.org/blog/ACL-2023-policy.

[19] Authorship [EB/OL]. [2023-02-09]. https://www.nature.com/nature/editorial-policies/authorship.

[20] Kirchenbauer J,Geiping J, Wen Y X,et al. A Watermark for Large Language Models[EB/OL]. [2023-02-22]. https://arxiv.org/abs/2301.10226.

[21] 警惕美国利用人工智能加剧对华认知战 [EB/OL]. [2023-02-20]. https://www.thepaper.cn/newsDetail_forward_21933225.

[22] OPWNAI: Cybercriminals Starting to Use Chatgpt [EB/OL]. [2023-02-09]. https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt.

[23] Liu J W,Kang Y Y,Tang D,et al. Order-Disorder:Imitation Adversarial Attacks for Black-Box Neural Ranking Models[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. Los Angeles,CA,USA. New York:ACM,2022:2025–2039.

[24] Westerlund M. The Emergence of Deepfake Technology:A Review[J]. Technology Innovation Management Review,2019,9(11):39-52.

[25] ChatGPT and Bing Give Problems: Emotional Breakdowns, Strange Answers and More[EB/OL]. [2023-02-22]. https://en.softonic.com/articles/chatgpt-bing-problems.

[26] Mitchell E,Lee Y,Khazatsky A,et al. DetectGPT: Zero-Shot Machine-Generated Text Detection Using Probability Curvature[EB/OL]. [2023-02-22]. https://arxiv.org/abs/2301.11305.

[27] Dou Y,Forbes M,Koncel-Kedziorski R,et al. Is GPT-3 Text Indistinguishable from Human Text? Scarecrow: A Framework for Scrutinizing Machine Text[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers) . Dublin,Ireland. Stroudsburg,PA,USA:Association for Computational Linguistics,2022:7250-7274.