

ChatGPT and Library Domain Trends: Post-Print

Authors: Zhang Jianing, Song Xigui

Date: 2024-04-03T00:00:00+00:00

Abstract

[Purpose/Significance] ChatGPT, including its upgraded versions GPT-4 and GPT-4 Turbo, addresses the challenge of reciprocal human-computer interaction through its powerful natural language processing capabilities, making it possible for traditional libraries to evolve into smart libraries and return to their fundamental role of knowledge services. [Method/Process] Based on the destiny logic of libraries in the computer age, this article employs retrospective and comparative methods to reveal the dilemmas of knowledge services in traditional libraries and the endogenous growth value of ChatGPT for library smart services. From the perspective of ChatGPT, library operations present a new mirror image: intelligent retrieval, with its exceptional radiating capability, covers all traditional service items, and the focus of information consultation, subject services, and reading promotion will undergo cluster displacement, with boundaries becoming blurred. The inherent copyright risks of ChatGPT and subsequent deep institutional copyright risks, including viewpoint plagiarism, content simulation, and data exposure, compel society to ensure security through legislation, third-party supervision, and local area network isolation. [Results/Conclusion] ChatGPT will trigger a historical migration and reorganization of traditional library service methods, but this does not necessarily lead libraries toward an existential crisis; rather, it makes it possible for libraries to return to their fundamental knowledge service role under the impetus of new technologies. Institutional mergers, staff reductions, and team reconstruction are difficult to avoid, but the process may be gentle and imperceptible. Library evaluation criteria will shift from borrowing volume, click rates, and browsing volume to the depth and performance of participation in teaching and research. Knowledge services will transform from ubiquitous and chaotic to clear and self-aware. The threat posed by digital books and electronic reading will be substantially reduced, and libraries will more easily confirm their position within the socio-historical constitution.

Full Text

Abstract

[Purpose/Significance] ChatGPT, including its upgraded versions GPT-4 and GPT-4 Turbo, solves the challenge of peer-to-peer response in human-computer interaction through powerful natural language processing capabilities, making it possible for traditional libraries to evolve into smart libraries and return to their fundamental role of knowledge services. **[Method/Process]** Based on the developmental logic of libraries in the computer age, this paper employs retrospective and comparative methods to reveal the dilemmas of traditional library knowledge services and the endogenous growth value of ChatGPT for library smart services. From the perspective of ChatGPT, library operations present a new mirror image: intelligent retrieval covers all traditional service items with strong radiating capacity, while the focus of information consultation, subject services, and reading promotion undergoes cluster displacement with increasingly blurred boundaries. The inherent copyright risks of ChatGPT, along with secondary deep institutional copyright risks, viewpoint plagiarism, content simulation, and data exposure, urge society to ensure security through legislation, third-party supervision, and local area network isolation. **[Results/Conclusion]** ChatGPT will trigger a historic migration and reorganization of traditional library service models, but this does not necessarily lead libraries toward a critical life threshold. Instead, it enables libraries to return to knowledge service fundamentals under the blessing of new technologies. Institutional consolidation, personnel streamlining, and team reconstruction are inevitable, but the process may be gentle and imperceptible. Library evaluation standards shift from borrowing volume, click rates, and page views to the depth and performance of teaching and research participation. Knowledge services transform from ubiquitous and chaotic to conscious and self-aware. The threat of digital books and electronic reading is substantially reduced, making it easier for libraries to confirm their position in social and historical development.

Keywords: ChatGPT; library; traditional service model; AIGC; field trend

1 Introduction

In November 2022, OpenAI launched the conversational chatbot ChatGPT, triggering worldwide shock across specific industries and domains. Using non-specialized language to issue instructions or pose questions, users can achieve human-computer interaction for information retrieval, code editing, and more. As a Reinforcement Learning from Human Feedback (RLHF) model, ChatGPT possesses self-correction capabilities, refuses unreasonable content, and demonstrates high self-expansion ability, approaching human-level information interpretation and output—a leap forward in human information processing. Libraries are institutions primarily focused on knowledge services. For a considerable period, they will need to address the opportunities and challenges brought by ChatGPT. This paper focuses on the field trends of libraries in the era of

strong artificial intelligence, exploring the path to industry leapfrogging and discussing how libraries can utilize ChatGPT-like technologies to reduce risks.

2 Technical Integration, Super-scenario Thinking, and Artificial Intelligence

ChatGPT (Chat Generative Pre-trained Transformer) has undergone two technological leaps since its inception. On March 15, 2023, the advanced version GPT-4 was released. Beyond quantitative differences such as supported text length and rate limit improvements, the most significant jump was the multi-modal mechanism that transcended single text status. GPT-4 can accept image and text inputs and process images through API interfaces. Similar to ChatGPT, GPT-4 represents a breakthrough achievement in pre-trained transformer models for artificial intelligence technology.

ChatGPT, powered by GPT-3.5, integrates the vast data of the rapidly developing 21st-century internet, using nearly 300 billion tokens and 175 billion parameters of corpus from diverse heterogeneous information sources. Its powerful intelligence is closely related to the massive, well-distributed corpus. While OpenAI has not disclosed corpus details as trade secrets, experts speculate it primarily used six types of digital materials including Wikipedia and academic journals [1]. The main corpus from the internet contains varying degrees of impurities, requiring data cleaning, filtering of low-quality pages, and deduplication.

GPT-4 Turbo's significance manifests in several aspects: text length jumped from GPT-4's initial 8,192 tokens to 128,000 tokens (equivalent to 300 pages of standard text); seed parameters and system_{fingerprint} response fields enable deterministic output for each request; and the corpus cutoff date advanced from September 2021 to April 2023.

The most 震撼性 advancement lies in requiring no programming—individuals can build GPT rooted in their own knowledge bases using natural language. After uploading to the GPT Store, creators can receive revenue shares based on user engagement. OpenAI has taken decisive steps on critical issues concerning the company's survival.

From a technical perspective, neither Transformer architecture, large language model pre-training, nor human feedback reinforcement learning models are OpenAI's original innovations. However, the non-linear superposition of multiple existing technologies has enabled OpenAI to stand atop the emergent capabilities of artificial intelligence technology. In the first round of AGI competition, OpenAI's undisputed victory stemmed from its general-purpose approach. However, the pioneering ChatGPT also faces significant challenges: its unidirectional autoregressive model cannot guarantee response accuracy, while Google's Bard and other tech giants' AI robots have obvious advantages in this regard. The emergence of GPT-4 Turbo has largely eliminated industry concerns, as OpenAI continues to lead in subsequent technological implementation.

3 Overview of ChatGPT and Library & Information Service Research

ChatGPT has existed for only one year, offering dazzling possibilities, and the industry remains in the verification stage of its known and unknown capabilities. Research on ChatGPT's application to library services is limited to a small number of articles. ADETAYO noted that ChatGPT AI robots can replace manual library services, provide website navigation as virtual assistants, and trigger personnel reduction in cataloging [6]. PANDA and KAUR compared ChatGPT with traditional library Q&A robots, pointing out that due to advantages in scalability and analytical capabilities, ChatGPT will become an important component of future library information services [7]. LUND and TING believe ChatGPT can effectively enhance library literature retrieval, metadata generation, and content creation services, though privacy issues and information bias require sufficient attention [8]. FERNANDEZ summarized key features of large deep learning models like ChatGPT and tested ChatGPT's capabilities in practical application scenarios such as reading promotion [9].

In domestic research, GUO Yajun et al. noted that ChatGPT catalyzes library smart service transformation with 4T characteristics: Transformation (smart service transformation), Translation (interactive language translation), Transmission (instant three-dimensional information dissemination), and Treatment (real-time task processing). Libraries need to improve service efficiency from technical and management paths to advance smart library construction [10]. CHU Jiewang et al. discussed the evolution of AI technology and its history with smart library construction, pointing out AIGC's practical significance for library information retrieval and special group services [11]. LI Shuning and LIU Yimeng evaluated ChatGPT from usability, system experience improvement, and other aspects, concluding that while ChatGPT significantly improves library service efficiency, intelligent consultation system construction, and knowledge services, it also impacts library resource organization and team building [12]. ZHANG Zhixiong et al. viewed ChatGPT as an opportunity to realize the AGI dream from the perspective of machine learning and unregulated pre-training [13].

4 The Fate of Libraries in the Computer Age and the Stimulating Effect of Strong AI ChatGPT

The fate of libraries in the computer age entered academic view as early as the 1970s-1980s. In 1978, American library and information expert F. W. Lancaster asserted in his article "Libraries and Librarians in an Age of Electronics" that "we are rapidly and inevitably moving toward a paperless information society." A few years later, he warned again that libraries mainly process machine-readable documents, and in 20 years, current libraries might completely disappear. On January 2, 2011, Sullivan from Alfred University Library published "Academic Library Autopsy Report, 2050," predicting the future demise of libraries from

several perspectives: collection books 无人问津, training becoming redundant, reference work disappearing, and librarians moving into IT departments.

Lancaster's warnings occurred during the initial period of computing, yet libraries did not end at the turn of the century as he predicted. Sullivan's prophecy, though, is clearly manifesting: electronic books have become irreversible, online access grows geometrically, and the proportion of electronic literature increasingly weighs heavier. While libraries continue developing literature retrieval education, the relative lag in retrieval models and limited technical power make it difficult to bridge the gap between readers and literature.

ChatGPT breaks the application scenario barriers of artificial intelligence, achieving the largest market 下沉 in history. Within one month of its English launch, it gained over 100 million users, becoming the fastest-growing internet application. After more than a decade of exploration, a consensus has gradually formed: smart libraries represent the inevitable stage of digital library development. However, limited by AI technology development levels, how data mining could universally apply to library knowledge services remained an unsolved problem. With ChatGPT's emergence, the problem suddenly becomes clear. ChatGPT's powerful information processing capabilities enable libraries to return to literature knowledge service fundamentals without needing ubiquitous services—the collection resources themselves suffice. This provides important development opportunities for libraries repeatedly impacted in the computer age.

5 Facilitating Endogenous Growth of Smart Services, Enabling Libraries to Return to Knowledge Service Fundamentals

5.1 Omniscient and Intelligent Literature Retrieval Strategies

Regardless of willingness, traditional libraries have long faced the contradiction between readers' unlimited demand for value-added knowledge services and libraries' near-zero supply capacity. Before ChatGPT, people might not have realized this issue. Library collections have already shifted from paper to paper-electronic dual modes. ChatGPT enables zero-threshold, unlimited satisfaction of reader needs. What previously required massive labor—customized writing, cross-temporal comparative demonstration, high-quality analysis, and blind-spot-free deep interaction—can now be easily achieved through ChatGPT. The quality and value of ChatGPT responses depend on the questions asked; good prompts elicit richer, more powerful responses. Librarians liberated from basic consultation can transform into intelligent retrieval specialists, researching and aggregating all effective retrieval solutions to guide readers infinitely closer to learning.

5.2 Information Consultation Shifting Toward Professionalism While Maintaining Necessary General Share

Reference consultation is an information transfer process between librarians and readers. Reader needs are decomposed by librarians into corresponding retrieval indicators based on collection external characteristics and content ontology. According to Megan Dempsey' s survey of the US National Library of Sciences, 62% of librarian responses belong to basic information and guidance acquisition, while the ratio of general to professional questions at the USF Tampa Library is 9:1 [20]. Although libraries are expanding services through networks—such as virtual reference, non-real-time online consultation, and customized information push [21]—in-depth information consultation remains a difficult problem in smart library construction. Only 18 of 42 “double first-class” university libraries surveyed by Zhong Huan et al. have established subject service platforms, with most suffering from outdated content, low update frequency, and difficult information retrieval [24].

ChatGPT, with its large natural language model core and 175 billion parameters trained on big data, possesses powerful computing capabilities. Combined with libraries' existing databases, it can effectively meet the needs of subject service platforms, breaking through constraints imposed by subject librarian team quality and individual professional direction. This drives the intelligent transformation of subject services from “subject librarian-centered” to “subject service platform-centered,” combining human resources with instantaneous, all-domain academic frontier push to achieve high-quality services.

5.3 Subject Services Advancing Toward Instantaneous, All-Domain Academic Frontiers

Subject services provide comprehensive information assurance for scientific research based on literature information resources. Traditional library subject services mainly focus on cultivating subject librarians, whose professional capabilities determine service quality and constitute the internal source of continuous innovation. However, high-quality subject services require both high-caliber librarian teams and high-level subject service platforms. ChatGPT' s built-in information induction capabilities can also reduce maintenance pressure on service platforms and communication processes, better serving discipline construction. The super computing power of ChatGPT, combined with libraries' existing databases, can effectively meet the needs of subject service platforms, breaking through constraints on subject service levels imposed by subject librarian team quality and individual professional direction.

5.4 Reading Promotion Achieving Precise Docking Between Collection Essence and Target Populations

The reading promotion concept emerged in 1997, relatively late but profoundly meaningful. UNESCO' s World Book and Copyright Day advocates “Reading

for All.” In short, reading promotion refers to activities conducted by social organizations or individuals to promote reading among people—extending the scope of reading’ s beneficial effects on individuals and society. Reading promotion transforms libraries from passive waiting to active outreach. The continuous enrichment of communication media cannot replace the original, creative presentation of massive literature, especially precious collection materials. How to accurately target reading promotion audiences has been an unsolvable puzzle since the concept’ s birth. In this regard, ChatGPT demonstrates extraordinary excellence. ChatGPT’ s introduction will achieve seamless docking between precious collection documents and target populations, including information retrieval disadvantaged groups, promoting integration between reading promotion and information consultation/subject services.

6 Risks and Strategies

6.1 Risks

As a cutting-edge AI achievement, ChatGPT can be introduced into library work as actual productive forces to improve quality and efficiency. However, ChatGPT’ s inherent limitations pose significant challenges, with the most prominent impact on education and research. American linguist Chomsky even called ChatGPT essentially a high-end plagiarism system that discovers patterns from massive data and strings them together like human writing [27]. ChatGPT exhibits notable deficiencies: (1) synonymous narration—paraphrasing internet content; (2) meaningless narration—providing logically compliant but practically meaningless replies; (3) information ambiguity—when questions are unclear or beyond response scope, ChatGPT may provide ambiguous or incorrect answers. The latter two points are most critical. The unidirectional autoregressive model determines that ChatGPT can only mechanically replicate existing corpus statements, unable to provide novel knowledge creation. This may cause factual errors for ordinary readers or interdisciplinary researchers lacking sufficient academic background.

6.1.1 Copyright Risks ChatGPT trains on publicly available internet data, and the copyright ownership of this data exists in a legal gray area. Cong Lixian and Li Yonglin note that ChatGPT generates content through algorithmic data recombination, creating corresponding copyright issues in three stages: data mining, content generation, and generated content usage [28]. These three stages all involve copyright risks. As ChatGPT itself admits, even though it tries to minimize copyright infringement, it cannot guarantee all used data has original authors’ permission. Pre-existing copyright risks and secondary deep institutional copyright risks 动摇 library academic coordinates. Several institutions including Sciences Po, University of Washington, and University of Hong Kong have explicitly prohibited students and faculty from using ChatGPT or similar AI to complete papers or assignments.

6.1.2 Academic Integrity ChatGPT is strong AI with human-equivalent intelligence and far-superior machine computing power. It can both assist researchers in saving information collection and trend analysis time, and organize text to form professionally competent articles. Since measuring ChatGPT's specific contribution proportion in research is difficult, this inevitably intensifies the possibility of improper profit-seeking through AI 代替 academic research. More critically, ChatGPT generates content through probabilistic distribution random matching, fitting human expression habits, which easily forms invisible academic integrity violations beyond subjective awareness—more undetectable and fatal than explicit plagiarism.

6.1.3 Data Security With GPT-4 Turbo's emergence, OpenAI has initiated its strong AI 落地 program. Organizations can build exclusive GPTs for internal use or create general-purpose GPTs based on their own knowledge bases [29]. For libraries, while accessing ChatGPT, how to maximize data security is the key issue. A feasible solution is establishing physical isolation between ChatGPT and enterprises, creating proprietary LANs unauthorized to outsiders. In fact, early ChatGPT operated within LANs [35]. Libraries, as non-profit organizations and important components of public cultural service systems, have public welfare as their highest goal. This positions libraries advantageously as public welfare third parties to participate in ChatGPT supervision, balancing commercial attributes and strengthening social responsibilities.

6.2 Strategies: Sharing Technology Dividends, Constructing Human-Machine Symbiotic Order

6.2.1 Libraries Participating in Supervision as Public Welfare Third Parties The emergence of ChatGPT has accelerated AI's penetration into all industries, with many still confused about management and usage. Libraries can participate in rule-making as public welfare third parties, establishing open and standardized application directions, reducing operational barriers, responding to user concerns, and improving data evidence systems. Libraries can also hold special authorizations to conduct AI-generated text forensics, helping academia and education certify actual learning capabilities and research contributions, preventing 违规 use of ChatGPT to circumvent plagiarism detection.

6.2.2 Establishing LANs to Ensure Data Security The fundamental factor determining ChatGPT's enterprise 落地 is vertical, personalized training. Under non-absolute physical security guarantees, establishing proprietary LANs is a realistic approach. University libraries possess massive knowledge integrations with strong property rights attributes, open only to specific populations. This determines that ensuring library knowledge resources do not expose themselves to the open internet is the primary issue in utilizing new AI technologies.

7 Conclusion

This wave of artificial intelligence triggered by ChatGPT will profoundly affect the library industry, changing operational models. On the basis of saving human resources and management costs, ChatGPT optimizes traditional library business types and improves service quality and efficiency. It will cause historic migration and reorganization of traditional library service methods, but this does not necessarily lead libraries toward a critical life threshold. Instead, it makes possible a return to knowledge service fundamentals under new technology blessings. Consolidation, personnel streamlining, and team reconstruction are inevitable, but the process may be gentle and imperceptible. Library quality standards and evaluation systems will fundamentally change, with core indicators shifting from borrowing volume and click rates to participation depth and performance in teaching and research processes. Knowledge services transform from ubiquitous and chaotic to conscious and self-aware. The threat of digital books and electronic reading is substantially reduced, making it easier for libraries to confirm their position in social and historical development. ChatGPT will inevitably change libraries, transforming them from holders of cultural and academic resources into active providers and deep processing exporters, bringing new opportunities to library education.

References

- [1] THOMPSON D A. ChatGPT 数据集之谜 [EB/OL]. (2023-02-15)[2023-03-15]. <https://zhuanlan.zhihu.com/p/606432878>.
- [6] ADETAYO A J. Artificial intelligence chatbots in academic libraries: The rise of ChatGPT[J]. Library hi tech news, 2023,40(3): 18-21.
- [7] PANDA S, KAUR N. Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers[J]. Library hi tech news, 2023, 40(3): 22-25.
- [8] LUND B, TING W. Chatting about ChatGPT: How may AI and GPT impact academia and libraries?[J]. SSRN electronic journal, 2023, 40(3): 26-29.
- [9] FERNANDEZ P. Through the looking glass: Envisioning new library technologies[J]. Library hi tech news, 2014, 31(7): 5-8.
- [10] GUO Y J, GUO Y R, LI S, et al. ChatGPT empowers library smart service: Characteristics, scenarios and realization paths[J]. Library development, 2023(2): 30-39, 78.
- [11] CHU J W, DU X X, LI J X. The impact and application prospects of artificial intelligence generated content on smart library services[J]. Information studies: Theory & application, 2023, 46(5): 6-13.
- [12] LI S N, LIU Y M. Opportunities and challenges for library from the rise of ChatGPT-like intelligent chat tools[J]. Library tribune, 2023, 43(5): 104-110.

- [13] ZHANG Z X, YU G H, LIU Y, et al. The influence of chat GPT on library & information services[J]. Data analysis and knowledge discovery, 2023, 7(3): 36-42.
- [14] ZHAO R X, HUANG Y W, MA W L, et al. Insights and reflections of the impact of ChatGPT on intelligent knowledge services in libraries[J]. Journal of library and information science in agriculture, 2023, 35(1): 29-38.
- [16] SULLIVAN T B. Academic library autopsy report, 2050[EB/OL]. (2011-01-11)[2023-06-25]. <http://simyjs.whu.edu.cn/xueshukeyan/xueshudongtai/2016-11-19/2643.html>.
- [17] AITTOLA M, RYHANEN T, OJALA T. Smart library: Location-aware mobile library service[C]//CHITTARO L. International Conference on Mobile Human-Computer Interaction. Berlin, Heidelberg: Springer, 2003: 411-416.
- [19] WU Y P, LI X D, AI C Y. Research on users behavior of resource discovery service - A case study of Peking university library[J]. Journal of academic libraries, 2023, 41(1): 70-77.
- [20] TANG Q. A study on tiered reference in libraries[J]. Library science research & work, 2022(8): 59-64, 69.
- [21] ZHONG H, WANG T Y, MA X F. A research on the status and optimization of subject service platform construction of “double first-class” university libraries[J]. Research on library science, 2021(21): 75-81.
- [23] YANG C, JIANG R B, XIONG L. Subject service of the second generation subject librarians[J]. Journal of library science, 2010, 32(10): 52-54.
- [24] CAO N, YANG Q. Preliminary study on the development ideas of reference and consulting services for smart libraries[J]. Journal of the national library of China, 2022, 31(3): 22-28.
- [25] WANG B. Research on the definitions of reading promotion and library reading promotion - Also on how to realize and learn fashionable cases of library reading promotion[J]. Library tribune, 2015, 35(10): 1-7.
- [26] NING X, LI Z, HUANG N. The path of library intelligent reading promotion service[J]. Journal of academic library and information science, 2022, 40(5): 38-42.
- [27] CHOMSKY N. The false promise of ChatGPT[EB/OL]. (2023-03-10)[2023-06-25]. https://www.sohu.com/a/652088066_{260616}.
- [28] CONG L X, LI Y L. Copyright risk of chatbot-generated content and its governance - From the perspective of ChatGPT application scenario[J]. China publishing journal, 2023(5): 16-21.
- [29] OpenAI GPT Store, GPT-4, API price 下调 nearly[EB/OL]. (2023-11-07)[2023-11-13]. <https://www.163.com/dy/article/IIUHNIJF0512B07B.html>.

[30] WANG Z Q, AYIBUEN B. Beyond ChatGPT: Possible value and traps of GPT-4 application in future university teaching[J]. Modern distance education, 2023(3): 3-11.

[31] 国家互联网信息办公室关于《生成式人工智能服务管理办法》公开征求意见的通知 [EB/OL]. (2023-04-11)[2023-05-10]. http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm.

[33] 人工智能的风险与监管完善 [J]. 上海市经济管理干部学院学报, 2022, 20(6): 29-38.

[35] 监管八项措 [EB/OL]. (2023-07-22)[2023-07-29]. http://news.cnr.cn/native/gd/20230722/t20230722_526340.htm.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.