The Reliability and Usability of ChatGPT for Library Metadata

Jenny Bodenhamer Assistant Professor, Digital Services Librarian Oklahoma State University

Abstract

At the end of November 2022, OpenAI launched ChatGPT, an artificial intelligence chatbot, and it quickly became a world-wide phenomenon. Instantly, it became a subject of controversy and concern as well as praise. Schoolteachers and professors grew worried as ChatGPT was used to create content for everything from high school assignments to scholarly works. Lazy writers aside, ChatGPT's output has often proved to be inaccurate to the point of complete fabrication. ChatGPT has also regularly misattributed the sources of its information, even giving the wrong author for large blocks of text.

With all ChatGPT's weaknesses, does ChatGPT have any beneficial uses for catalogers and metadata professionals? As a field, information professions are regularly challenged to do more work, more accurately, in less time. Does ChatGPT offer any reliable, accurate services at this time to assist these professionals in completing their tasks?

This study seeks to evaluate the weaknesses and strengths of ChatGPT as it tries to perform three common cataloging/metadata tasks: 1) assigning classification numbers, 2) choosing Library of Congress subjects headings, and 3) harvesting keywords. Over the course of four months, it will ask ChatGPT a standardized list of questions on these topics. Then it will collate and evaluate ChatGPT's performance. In the end, this study will offer its findings as well as best practices for using ChatGPT in cataloging and metadata tasks.

The Reliability and Usability of ChatGPT for Library Metadata

Introduction

At the end of November 2022, OpenAI, an artificial intelligence research laboratory, released the product ChatGPT to the public. It quickly became the fastest growing app in the history of web applications (Gordon, 2023). It had over one million users in 5 days. Since then, the world has begun to see a revolution.

ChatGPT's origins began when a group of investors and computer scientists started OpenAI in 2015. Their goal was to create a generative artificial intelligence that could work in the background of other applications such as video games and websites. This became the series of "Generative pre-trained transformer" models, i.e., GPT-1. GPT-2, GPT-3, and GPT 4.

To explain, GPT software would work like a human brain. Using the research of scholars like Geoffrey Hinton, the software would mimic a "neural network" (Metz, 2023, May 1). This type of neural-network-based software would draw on complex mathematics and statistics in order to mimic neurons in the human brain. Once this framework was in place, the software ingested enormous amounts of data and processed that data through pattern recognition. In essence, GPT was learning. Pattern recognition performance is better the more data that the system consumes; so, the creators of GPT fed this software a large amount of text both curated and from the internet including Google, Google Scholar, and more (OpenAI, 2023). At this point, GPT could generate new text back to the user based on past patterns. One word at a time, GPT predicts the next word in a series based on these patterns until GPT finishes the task. This is an example of generative artificial intelligence. Since it uses a large amount of text, it is also called a large language model (LLM) (What is ChatGPT, n.d.) (Heaven, 2023) (Cox & Tzoc, 2023).

When OpenAI released ChatGPT to the public in November 2022, it became the first way the public could access GPT software directly, through a chat box. OpenAI had few expectations for their product. Jan Leike, the leader of OpenAi's alignment team, mentioned in an interview, "It's been overwhelming, honestly. We've been surprised, and we've been trying to catch up" (Heaven, 2023). However, since ChatGPT was so accessible and user-friendly, the users have used it in diverse and creative ways. In academia, there are a broad range of uses for ChatGPT from students plagiarizing assignments to science departments synthesizing large data sets.

ChatGPT has remarkable abilities to quickly process copious amounts of data and generate information from that data for its users. Unfortunately, its very strengths are its greatest weaknesses. For example, ChatGPT can quickly find the most common answer to a question, but if the information ChatGPT is accessing is wrong, then ChatGPT will be wrong as well. It is only as smart as the humans entering information onto the internet. Also, many reliable sources are only available behind paywalls that ChatGPT cannot access. Despite these limitations, the creators of ChatGPT can create new parameters and upgrades to help ChatGPT analyze and retrieve the most reliable sources of information even when a majority of the information available is inaccurate. In fact, ChatGPT has already had at least eight releases (*ChatGPT*—*Release Notes*, n.d.). But even then, there is a possibility for bias if the programmers create parameters in a biased way.

Another common strength/weakness of ChatGPT is how its "brain" works through the process of pattern recognition. Alkaissi and McFarlan explain ChatGPT is a "natural language generator." Through the rules of language and pattern recognition, Chat GPT provides information one word after another into a response. In this way, ChatGPT can pull together many sources of information, but sometimes it will create a pattern that does not exist i.e., it hallucinates (Alkaissi & McFarlane, 2023). For example, ChatGPT will provide a reference for topic that looks completely logical with a notable author and publisher, but the title will not actually exist, and any links to that article may go to a different article. In contrast on a different occasion, ChatGPT may provide a reference that is perfectly correct.

Thesis

With this understanding of ChatGPT, can librarians utilize ChatGPT to perform metadata tasks in their libraries? How can ChatGPT's strengths assist catalogers and metadata librarians in their work? What are the pitfalls of using ChatGPT to complete metadata tasks, and can information professionals avoid these pitfalls? Over the course of the last four months, the author has asked ChatGPT a standardized list of questions on metadata topics and has collated snapshots of that data over time. Using this bank of information, this study evaluates the weaknesses and strengths of ChatGPT as it tries to perform three common cataloging/metadata tasks: 1) assigning classification numbers, 2) choosing Library of Congress subjects headings, and 3) suggesting keywords.

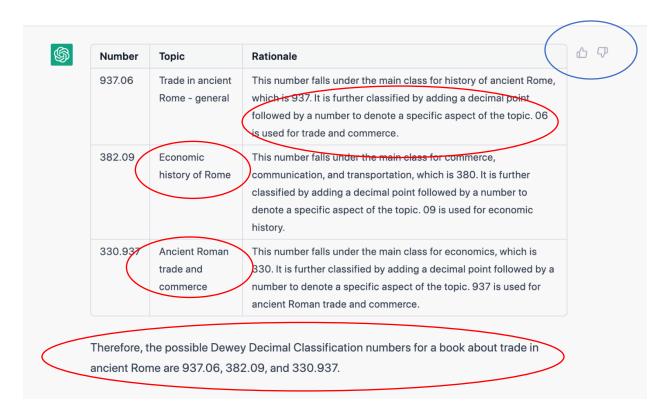
Classification

In early March 2023, the author received an email from Clarke Iakovakis, Scholarly Services Librarian at Oklahoma State University, wondering how well ChatGPT did at assigning Dewey Decimal Classification (DDC) numbers to a book about trade in ancient Rome (C. Iakovakis, personal communication, March 13, 2023). This was an interesting question with multiple aspects to consider. First, the quality of ChatGPT highly depends on the type of prompt the user asks. The process of asking for a useful, detailed prompt is called prompt engineering. In fact, the more details the user supplies the better the answer. Several professionals have begun to teach prompt engineering like Dave Birss and Steve Hargadon (Birss, 2023) (Hargadon 2023). They encourage being polite to ChatGPT and to specify an audience, length, and context for a dependable answer. If the user does not exactly receive the answer they were seeking, they can modify that answer with a follow-up question. In this case, the colleague formed his prompt carefully and asked ChatGPT to give the information in the form of a chart with a labeled topic and descriptive rationale.

Figure 1

ChatGPT Prompt and Dewey Decimal Classification Table from March 13, 2023

provide three dewey decimal classification numbers down to the thousandths place for a book about trade in ancient rome? provide the results in a table with the headings: number, topic, rationale



At first glance, the answer by ChatGPT looks accurate and professional. The chart is easy to read, but while the rationale sounds convincing, it has errors that a professional cataloger would observe (see red circles). ChatGPT's first error is the line "06 is used for trade and commerce." The "06" in this context identifies a period of ancient Roman history, specifically the Roman Empire from 31 B.C.E. to 476 C.E. The second error happens in the topic heading for the Dewey number 382.09. The rationale is mostly correct, but WebDewey, the authoritative source for DDC numbers, would describe this topic closer to the "History of International commerce (Foreign trade)" (*WebDewey*, n.d.). The third row has a rationale that is almost correct. A more appropriate label for the topic of 330.937 would be the "economics of ancient Rome." The last issue in this response is when ChatGPT creates a conclusion that claims that these are *the possible* DDC numbers for a book about trade in ancient Rome. There could indeed be other possibilities. It was the prompt that asked for three examples.

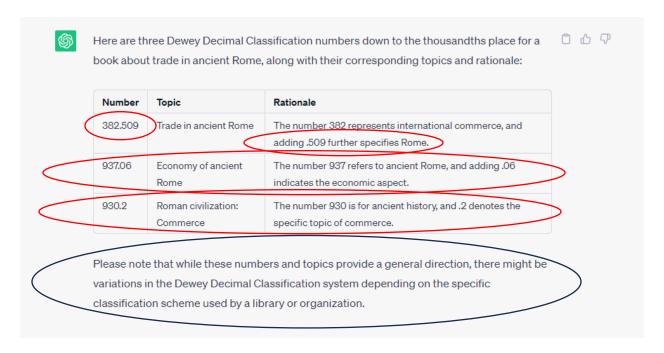
On the positive side, none of these call numbers are completely wrong. A professional cataloger might use 937.06 or 330.937 depending on their collection. The author (without ChatGPT)

would have likely chosen 937.06 because there are many items on the Roman Empire in that section of the library at her university. However, at a university with a larger classics collection, a different cataloger might have chosen the Dewey number, 381.0937. Using WebDewey, 381 is the Dewey number for trade, and the .093706 specifies ancient Rome during the time of the empire (*WebDewey*, n.d.). Unfortunately, ChatGPT makes its mistakes with such an authoritative voice that a novice might not question ChatGPT's response and accept it as accurate.

When a user recognizes that ChatGPT has made a mistake, there is a feedback system. See the blue circle in Figure 1. The user may click the thumbs up or thumbs down and then provide additional feedback in a popup window. (For this study, the author did not use this feedback method with ChatGPT because she wanted to see how ChatGPT changed over time without her intervention.)

Figure 2

ChatGPT Prompt and Dewey Decimal Classification Table from June 1, 2023



The author used the same prompt again on June 1, 2023. See Figure 2. ChatGPT performed worse on its task than before. On the first row, ".509" does not specify Rome. The DDC number "382.5" is for import trade and "09" refers to the history of a topic. The second row makes the same mistake as it did in March. The third row does not even exist in the current Dewey Decimal System (*WebDewey*, n.d.).

There was one improvement. In the blue circle of Figure 2, ChatGPT creates a much better note describing its limits. This note does not indicate that ChatGPT might have difficulty assigning Dewey numbers; however, it does warn the user to evaluate its answer.

Users of ChatGPT can understand these weaknesses when they remember how both Dewey and ChatGPT work. WebDewey is a subscription-based resource. ChatGPT cannot directly access

WebDewey and its content without permission. ChatGPT is trying to create DDC numbers based on what is widely available on the Internet. Often it cannot even access the information housed on library catalogs unless that catalog is integrated with a search engine. A programmer could train ChatGPT to create better DDC numbers, especially if ChatGPT had access to data about the Dewey Decimal System. ChatGPT also may have made mistakes because of its pattern recognition capabilities. It is possible ChatGPT chose these answers because they were either more numerous than other data points, or ChatGPT could not determine a pattern in the data that it had to create DDC numbers accurately.

Library of Congress Subject Headings (LCSH)

DDC numbers were behind a paywall. Library of Congress Subject Headings (LCSH) are not. They are readily available online. And while one cannot use a search engine to find Library of Congress catalog records directly, anyone can find the Library of Congress catalog online and then search for individual records. There are no legal/monetary barriers to access. However, the author does not know if ChatGPT has access to these records and to what extent.

Last year, Charlene Chou and Tony Chu used BERT (Bidirectional Encoder Representations from Transformers), a product similar GPT but by Google, to assign LCSHs to metadata records in Project Gutenberg, an e-book collection. Their hope was to automate the subject indexing for this and other digital library collections that were missing subject information. The initial set up for using BERT was complex, and the average librarian would find it difficult to replicate. However, the project was successful, and BERT created usable LCSH for over 95% of the records they tested (Chou & Chu, 2022).

BERT was able to reliably choose LCSH. Could ChatGPT assign subject headings as well? In April 2023, the author expanded the study to include the analysis of ChatGPT's ability to choose Library of Congress Subject Headings. Maintaining the topic of trade in ancient Rome, the author chose the work *Trade-routes and commerce of the Roman Empire* by M. P. Charlesworth as a baseline for ChatGPT performance, and ChatGPT performed considerably better at this task than with DDC numbers.

Over decades, catalogers at the Library of Congress have created four records for different editions of the book mentioned above. The most complete record that they created was https://lccn.loc.gov/79118637. It has three LCSHs.

- 1. 651 0 Rome -- Commerce -- History.
- 2. 650 0 Commerce -- History -- To 500.
- 3. 651 0 Rome -- Economic conditions -- 30 B.C.-476 A.D.

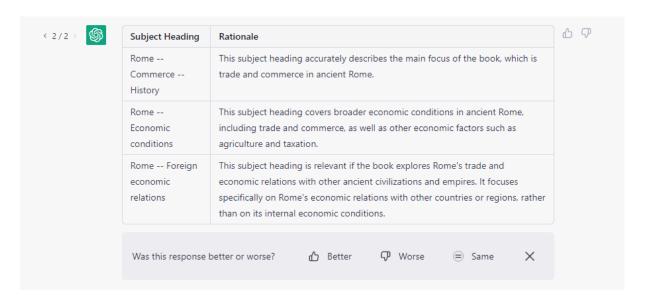
The three other Library of Congress records for this book had the LCSHs:

- 1. 651 0 Rome -- Commerce -- History.
- 2. 650 0 Commerce -- History.
- 3. 651 0 Rome -- Economic conditions.

The author gave ChatGPT the prompt "Provide three possible Library of Congress subject headings for a book about trade in ancient Rome. Provide the results in a table with the headings: subject heading, rationale." With this prompt, ChatGPT suggested almost the same LCSHs as the Library of Congress records.

Figure 3.

ChatGPT Results for the Prompt "Provide three possible Library of Congress subject headings for a book about trade in ancient Rome. Provide the results in a table with the headings: subject heading, rationale." (April 3, 2023)

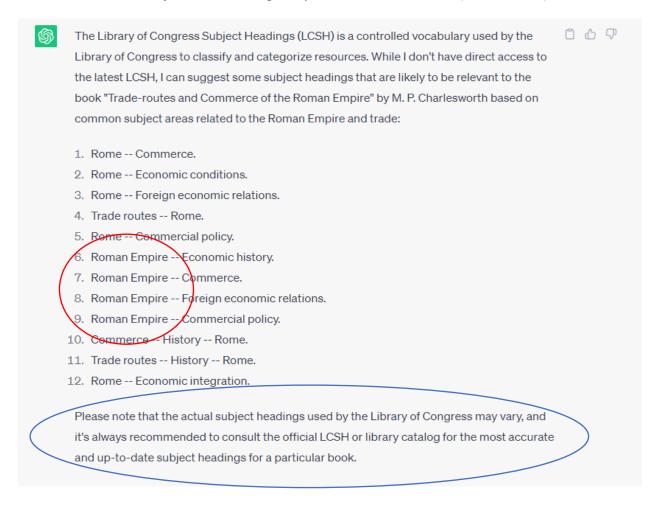


ChatGPT supplied two of the exact LCSHs as the group of three Library of Congress records, and it supplied an equally valid subject heading that Library of Congress did not include. In fact, since ChatGPT did not have more description in the prompt, it would not have been able to create more detailed LCSHs than it provided.

Later in June 2023, the author asked ChatGPT to complete a slightly different prompt, "Assign Library of Congress Subject Headings to Trade-routes and commerce of the Roman Empire by M. P. Charlesworth." In this case, ChatGPT provided more options but not all of them were authorities maintained by the Library of Congress.

Figure 4

ChatGPT Response to the Prompt "Assign Library of Congress Subject Headings to Traderoutes and commerce of the Roman Empire by M. P. Charlesworth." (June 1, 2023)



In this case, the Roman Empire is not a valid LCSH. Rome is the correct geographical authority. The first three subject headings that ChatGPT provided are correct. The use of the subdivision "Commercial policy" in this situation (instead of work of political science) is unusual, but the subdivision is not technically incorrect. With the accuracy of these results, the author would be comfortable recommending a new employee try ChatGPT for possible LCSHs; however, she would want novices to double check ChatGPT's responses to validate that they are a standard LCSH.

The June version added value to the April version by including an accurate disclaimer. These disclaimers have become more common in the last two months, and they make ChatGPT more transparent. It encourages users to check sources and evaluate ChatGPT's answers.

Keywords

A keyword search is substantially different from a Library of Congress Subject Heading search. With LCSHs, professionals create a controlled term and then they assign them methodically to a metadata record such as a MARC record. In contrast, a keyword search is usually a full text search of the Internet, a database, or a website. The most successful keyword searches are those in which a user chooses keywords that match the same vocabulary that an information provider supplies. Sometimes, a user must try several synonyms to find the results that they want in a search. It is in the best interest of both users and providers to know the most common keywords for a topic. Since ChatGPT's strength is in the amount of data it has processed and its pattern recognition, Chat-GPT should be an excellent source for determining useful keywords (Keyword - Definition and Meaning, n.d.).

The author prompted ChatGPT to provide a list of keywords in both April 2023 and June 2023 and gained the following feedback.

Figure 5

ChatGPT Response to the Prompt "Assign Keywords to Trade-routes and Commerce of the Roman Empire by M. P. Charlesworth." (April 7, 2023)

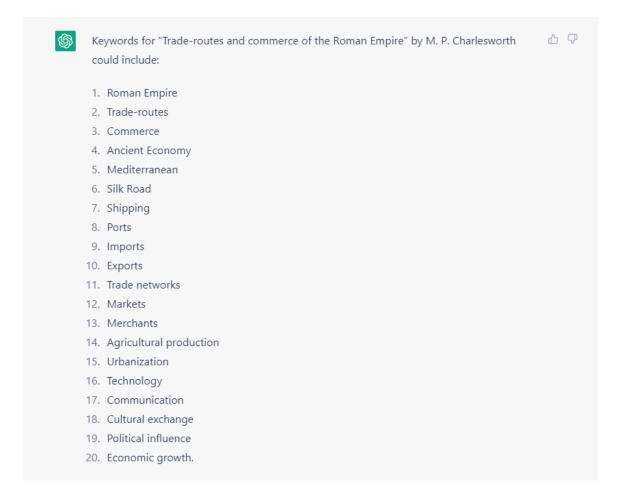


Figure 6

ChatGPT Response to the Prompt "Assign Keyword to Trade-routes and Commerce of the Roman Empire by M. P. Charlesworth." (June 1, 2023)



Here are some keywords that can be associated with the trade routes and commerce of the Roman Empire as discussed in M. P. Charlesworth's work:





- Silk Road: The Silk Road was a network of ancient trade routes that connected the East and West, facilitating the exchange of goods and ideas between the Roman Empire and regions such as China and India.
- Mediterranean Sea: The Mediterranean Sea played a crucial role in Roman trade, as it served as a major transportation route for goods and connected various regions within the empire.
- Ports: Roman ports, such as Ostia and Portus, were important hubs for maritime trade.
 These ports facilitated the movement of goods and were centers of economic activity.
- Commerce: Commerce refers to the buying and selling of goods, which was a vital aspect
 of the Roman economy. It involved various activities, such as trade, manufacturing, and
 financial transactions.
- Import/Export: The Roman Empire engaged in both imports and exports. They imported luxury goods like silk, spices, and precious metals from the East, while exporting goods such as grain, wine, olive oil, and pottery.
- Coinage: Roman coins, such as the denarius, served as a medium of exchange in commercial transactions. The stability and widespread use of Roman currency facilitated trade and commerce.
- Trade networks: The Roman Empire developed extensive trade networks that spanned across Europe, Africa, and Asia. These networks connected various regions and facilitated the movement of goods and cultural exchange.
- 8. Trade routes: Roman trade routes, both overland and maritime, connected different provinces and regions. These routes included the Via Appia, Via Domitia, Via Augusta, and the Egyptian Red Sea trade routes.
- Economic growth: The Roman Empire experienced significant economic growth due to its extensive trade and commerce networks. This growth contributed to the prosperity and stability of the empire.
- 10. Commercial hubs: Certain cities, such as Rome, Alexandria, Carthage, and Ephesus, served as major commercial hubs within the Roman Empire. These cities attracted merchants and traders, and their markets were centers of economic activity.
- 11. Trade goods: Various goods were traded within the Roman Empire, including agricultural products, textiles, spices, luxury items, metals, and pottery. These goods were exchanged both domestically and internationally.
- 12. Roman merchant class: The Roman merchant class played a crucial role in trade and commerce. They were involved in long-distance trade, managed commercial enterprises, and contributed to the economic prosperity of the empire.

These keywords provide an overview of the trade routes and commerce within the Roman Empire and can be used as a starting point for further exploration of M. P. Charlesworth's work. Both of ChatGPT's responses would be especially useful for a student or other novice beginning to research a topic. In these cases, there are no wrong or right answers. ChatGPT finds common keywords for a topic according to the specifications of the prompt, and an information provider and/or an information seeker can use the keywords for their project. In fact, if any of these keywords did not work well, a person could ask ChatGPT to find more keywords or synonyms to the original words that it chose.

However, June's response is much more detailed than April's, and it includes some context and a useful disclaimer is its response. The author would encourage the use of ChatGPT for choosing keywords. It is an excellent use of ChatGPT's strengths with little sign of weakness.

Conclusion

Overall, in these examples, ChatGPT's performance has become more transparent and more complete in just four short months. ChatGPT showed considerable development, and this study only used the version of ChatGPT available with free access. ChatGPT Plus was released in March, and its subscription begins at \$20 a month. ChatGPT Plus is vastly more sophisticated because it uses GPT-4 instead of GPT-3 (Rogers, n.d.). The creators of GPT-4 are touting it to be a true *artificial general intelligence*. In other words, its creators are hypothesizing that GPT-4 is acting more like humans and animals and that it has some form of intelligent agency (Kestenbaum, 2023, June 24) (Bubeck et al., 2023). Nevertheless, ChatGPT will continue to improve and become more accurate, both the free version and the subscription version.

ChatGPT is a new resource. ChatGPT is an imperfect resource. But ChatGPT is a *learning* resource. ChatGPT is learning from its own mistakes. It has motivated creators who are striving to make it better. ChatGPT is taking feedback from its users and improving its responses. Finally, ChatGPT/ChatGPT Plus will harness bigger sets of data with each new version, and it will be able to analyze and produce even better results.

At the current time, catalogers and metadata librarians need to be cautious using ChatGPT in their tasks, especially in tasks that require subscription access such as the Dewey Decimal Classification System. As librarians, it is important to evaluate our sources. However, the usefulness of ChatGPT is improving quickly, and ChatGPT is already able to assist in choosing LCSH. It was also very helpful in selecting possible keywords for an internet search. Going forward, classification, subject headings, and keywords are just the beginning. Future areas of exploration could include creating abstracts and summaries. Catalogers and metadata librarians are necessary for the production of reliable and usable metadata, but they now can work with ChatGPT to create timely, authoritative, reliable records with excellent discoverability.

References

- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, *15*(2), e35179. https://doi.org/10.7759/cureus.35179
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838
- Birss, D. (2023). How to research and write using generative A.I. tools. [Video]. *LinkedIn Learning*. https://www.linkedin.com/learning/how-to-research-and-write-using-generative-ai-tools/how-to-write-an-effective-prompt-for-ai?autoSkip=true&resume=false&u=106104604
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. http://arxiv.org/abs/2303.12712
- *ChatGPT Release notes*. (n.d.). Retrieved June 29, 2023, from https://help.openai.com/en/articles/6825453-chatgpt-release-notes
- Chou, C., & Chu, T. (2022). An analysis of BERT (NLP) for assisted subject indexing for Project Gutenberg. *Cataloging & Classification Quarterly*, 60(8), 807–835. https://doi.org/10.1080/01639374.2022.2138666
- Cox, C., & Tzoc, E. (2023). ChatGPT: Implications for academic libraries. *College & Research Libraries News*, 84(3), 99–102. https://doi.org/10.5860/crln.84.3.99
- Gordon, C. (2023). ChatGPT is the fastest growing app in the history of web applications. *Forbes*. Retrieved June 29, 2023, from https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/
- Hargadon, S. (2023). Library 2.0's ChatGPT bootcamp for libraries and librarians. [Video]. *Library* 2.0. https://www.library20.com/bootcamps/chatgpt
- Heaven, W. D. (2023). The inside story of how ChatGPT was built from the people who made it. *MIT Technology Review.Com*. https://www.proquest.com/docview/2781894054/citation/8FB6B2E868924ED8PQ/1
- Kestenbaum, D. (2023, June 24). First contact. [Radio episode]. *This American Life*. https://www.thisamericanlife.org/803/greetings-people-of-earth/act-one-4
- *Keyword—Definition and meaning.* (n.d.). *Market Business News*. Retrieved June 29, 2023, from https://marketbusinessnews.com/financial-glossary/keyword/

- LaFountain, C. (2023). AI tools for libraries. Computers in Libraries, 43(4), 14–16.
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3), 26–29. https://doi.org/10.1108/LHTN-01-2023-0009
- Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. https://doi.org/10.1002/asi.24750
- Metz, C. (2023, May 1). 'The Godfather of A.I.' leaves Google and warns of danger ahead. *NYTimes.Com Feed*.

 http://global.factiva.com/redir/default.aspx?P=sa&an=NYTFEED020230501ej51001jy&cat=a&ep=ASE
- OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat
- Pophal, L. (2023). ChatGPT: Opportunities and risks related to AI-generated content. *Information Today*, 40(2), 36–38.
- Rogers, R. (n.d.). I finally bought a ChatGPT Plus subscription—And it's worth it. *Wired*. Retrieved June 29, 2023, from https://www.wired.com/story/chatgpt-plus-web-browsing-openai/
- Schreur, P. E. (2020). The use of linked data and artificial intelligence as key elements in the transformation of technical services. *Cataloging & Classification Quarterly*, 58(5), 473–485. https://doi.org/10.1080/01639374.2020.1772434
- WebDewey. (n.d.). Retrieved June 29, 2023, from https://dewey.org/webdewey/standardSearch.html
- What's the next word in large language models? (2023). Nature Machine Intelligence, 5(4), 331–332. https://doi.org/10.1038/s42256-023-00655-z
- What is ChatGPT? (n.d.). Retrieved June 27, 2023, from https://help.openai.com/en/articles/6783457-what-is-chatgpt