

# Generative AI Meets Cataloging Practice

## Findings from a Comparative Pilot Study

Greta Heng, Patricia Lampron, and Myung-Ja Han

---

### ABSTRACT

*This study evaluates the performance of four generative AI models—ChatGPT, DeepSeek, Gemini, and Copilot—in generating descriptive metadata for bibliographic resources. Models were tested on a small, diverse set of resources using four prompt types: a basic prompt, a basic prompt with an example, a detailed prompt referencing Resource Description and Access (RDA) guidelines, and a detailed prompt with an example. Results show that both detailed RDA guidance and the inclusion of sample outputs improved metadata quality, particularly in formatting and field structure. While DeepSeek and ChatGPT showed better performance on the tasks, all models displayed limitations in parsing and following the prompts, using descriptive metadata fields, analyzing subject headings, and assigning URLs. These findings suggest that while generative AI holds potential to assist in metadata creation, its current capabilities fall short of meeting cataloging standards without human review.*

### INTRODUCTION

There has been growing interest in applying artificial intelligence (AI) tools to cataloging and metadata creation. While conversations have explored both opportunities and challenges of integrating AI into library workflows, there is still a limited comparative understanding of how they function in real-world cataloging contexts.

This paper outlines a project that seeks to evaluate the practical capabilities of a selected group of generative AI models for descriptive metadata creation work. Focusing on a small set of resources including four English language monographs, one serial, one series, and one non-Latin-script title, we examine which aspects of cataloging these tools handle effectively and where they fall short in real-world metadata workflows. By comparing their outputs and identifying strengths, limitations, and potential areas of improvement, we aim to share our experience and provide catalogers with clearer guidance on how generative AI can be thoughtfully incorporated into cataloging workflows. The findings will support informed decision-making, help frame realistic expectations, and contribute to broader conversations about automation, labor, and quality in metadata creation.

### RELATED WORK

Research on the application of AI in cataloging and metadata creation can generally be divided into three areas: (1) the development of AI tools designed for cataloging tasks; (2) the testing and evaluation of specific AI tools for metadata creation, and (3) broad environmental scans that assess trends, opportunities, and challenges in the field.

#### *About the Authors*

**Greta Heng** ([gheng@sdsu.edu](mailto:gheng@sdsu.edu); corresponding author) is Cataloging and Metadata Strategies Librarian, San Diego State University. **Patricia Lampron** ([plampron@uci.edu](mailto:plampron@uci.edu)) is Cataloging and Metadata Librarian, University of California, Irvine. **Myung-Ja Han** ([mhan3@illinois.edu](mailto:mhan3@illinois.edu)) is Andrew Turyn Professor/Metadata Librarian, University of Illinois at Urbana-Champaign. © 2026.

Submitted: 18 August 2025. Accepted for Publication: 21 November 2025. Published: 15 June 2026.

The application of AI in cataloging and metadata tasks is not a new development within the profession, particularly in the area of subject indexing, which remains one of the most time-consuming yet critical processes for supporting subject-based access services. Effective subject indexing requires both expertise in the subject area and familiarity with the structure and application of controlled vocabularies. A notable example of AI applied to this domain is Annif, an automated subject indexing tool developed by Suominen.<sup>1</sup> Annif employs multiple algorithms and was trained using existing subject headings and metadata, enabling it to generate subject access points for new content in certain areas and formats. Golub et al. explored the challenges and potential benefits of (semi-)automated subject indexing, noting its potential to improve information retrieval systems.<sup>2</sup> Building on earlier work, Suominen, Inkinen, and Lehtinen introduced Pinto, a REST API service that offers a streamlined implementation of Annif's functionality.<sup>3</sup> Further developments and real-world implementation were documented by Golub et al., who presented a case study on the adoption of Annif by Swedish libraries for union cataloging.<sup>4</sup> Their findings highlighted both the challenges encountered and the substantial potential of AI-driven approaches to automating subject indexing at scale.

In addition to tool development, the testing and evaluation of AI tools for cataloging and metadata creation have received increasing attention, particularly over the last five years, especially in subject headings. These studies have varied widely in the AI tools and datasets used. Chou and Chu tested Bidirectional Encoder Representations from Transformers (BERT) models for the assignment of Library of Congress Subject Headings (LCSH) to items from the Project Gutenberg collection.<sup>5</sup> Their results demonstrated that the tool could generate relevant subject headings; however, they also noted that training AI models for subject indexing is a complex and time-consuming endeavor. Bodenhamer also tested ChatGPT on subjects, such as classification numbers, subject headings, and keywords.<sup>6</sup> Bodenhamer's research found that the performance of ChatGPT improved from March 2023 to June 2023 and speculated that it could be used for other areas of metadata. Taniguchi explored the ability of ChatGPT to create bibliographic records following the Resource Description and Access (RDA) standard.<sup>7</sup> The study found that while ChatGPT was able to follow RDA rules in many cases, it produced errors and inconsistencies, suggesting that it should be used primarily as an assistive tool for human catalogers rather than as an independent system for bibliographic record creation. Chow, Kao, and Li's study tested ChatGPT for assigning subject headings to electronic theses and dissertations using titles and abstracts.<sup>8</sup> The study found that while the output machine-readable cataloging (MARC) coding was mostly correct with subdivisions when evaluated against MARC guidelines, the quality of the subject headings was questionable, especially in accuracy and use of subdivisions. Similarly, Dobreski and Hastings evaluated the capabilities of AI chatbots—ChatGPT, Gemini, and Copilot—in assigning Dewey Decimal Classification numbers, Library of Congress Classification numbers, and LCSH.<sup>9</sup> They concluded that although these AI models could produce subject terms, the outputs were often overly general, inaccurate, or inconsistent, reaffirming the continued value of professional cataloging expertise.

Another stream of research is on the general application of AI in cataloging. Brzustowicz showcased how AI language models could impact the creation of cataloging records.<sup>10</sup> Using ChatGPT, the research found that “the model can extract essential metadata” and “generate MARC records” for resources in various formats and languages.<sup>11</sup> The paper stressed that while ChatGPT “has been trained on data from various catalogs, including OCLC's WorldCat,” the quality and amount of trained data may still not be enough to generate consistent quality catalog records.<sup>12</sup> The journal later published three responses to the paper, each highlighting inaccuracies in the

records produced by ChatGPT. DeZelar-Tiedman, Amram et al., and Floyd stressed that the records generated have significant errors and inaccuracies, and that it is unreliable and does not understand cataloging rules.<sup>13</sup> Moulaison-Sandy and Cable looked at the performance of generative AI in cataloging tasks and stressed the importance of the prompt used and that an understanding of how the tool should be used for specific tasks is critical in working with generative AI.<sup>14</sup> Zavalin and Zavalina compared the quality, accuracy, and completeness of metadata generated by three different generative AI tools. They found that the overall quality of the metadata was not ideal and required “involvement of human metadata experts.”<sup>15</sup>

Library systems vendors, most notably Ex Libris, EBSCO, and OCLC, have been actively testing and implementing AI-enhanced metadata editors within their platforms, as well as integrating AI into their discovery services. Ex Libris integrated AI into its Alma metadata editor and invited early adopters to begin testing in August 2024. According to a 2025 presentation by Dale et al., the AI performed well for fields requiring transcription, such as title, publication information, table of contents, and summary.<sup>16</sup> However, subject headings and authorized access points often contained inaccuracies, indicating a continued need for cataloger review and intervention.

Building on these previous studies, this paper aims to test four generative AI tools by evaluating their performance in cataloging and metadata creation tasks. Unlike earlier research that primarily focused on unveiling each AI tool’s general capabilities, this study examines how AI tools generate and structure bibliographic metadata based on scanned images of resources. Two base prompting strategies were used: one employing a basic prompt and the other with a detailed prompt that included additional instructions referencing RDA guidelines. Each prompt was tested under two conditions: a zero-shot setting (without an example) and a one-shot setting (with a single example provided). This design allowed for a more nuanced understanding of how prompt complexity and examples influence the selected generative AI’s ability to generate accurate and structured bibliographic information.

## METHODS

For these experiments, we selected three resource types: single-book monographs, monograph series, and serials, as they often contain different kinds of bibliographic information shaped by specific cataloging needs. Our test set included an English juvenile fiction monograph (Monograph 1: *Hitty Her First Hundred Years* by Rachel Field), an English adult fiction monograph (Monograph 2: *A Farewell to Arms* by Ernest Hemingway), an English nonfiction monograph (Monograph 3: *Catalog of Lincolniana* published by LeClaire Book Exchange), an English fiction monograph with a non-standard font (Monograph 4: *Gods’ Man* by Lynd Ward), an English monographic series (Series: *An Analysis and Correlation of Aircraft Wave Drag* by Roy V., Jr. Harris), an English serial (Serial: *1910 Annual Report of The Ontario Historical Society* published by The Ontario Historical Society), and one monograph in non-English language, Panjabi (Non-Latin: *Sandarbh Sewa* by Gursharan Kaur Brar).<sup>17</sup> For each item, we prepared a scanned image of the title page, and when the title page lacked sufficient information, such as date of publication, we included the title verso or a table of contents page to supplement the input.<sup>18</sup>

We selected four generative AI models for this study that are widely recognized and frequently cited in recent studies: ChatGPT 4o, DeepSeek R1, Google Gemini 2.5, and Microsoft 365 Copilot (Business Version). The project used default configurations of these models without any additional training. To evaluate their performance in bibliographic information generation tasks, we designed four prompts:

- Prompt 1: a basic prompt with minimal instruction,
- Prompt 1e: Prompt 1 with an example,
- Prompt 2: a detailed prompt referencing original RDA guidelines, and
- Prompt 2e: Prompt 2 with an example.

The two prompts and an example used for the tests are available in the Appendix.

Given the importance of RDA in cataloging, relevant detailed guidance based on RDA rules was incorporated into the design of Prompts 2 and 2e. The example record was used to assess whether providing a sample output would enhance the generative AI's comprehension of the task and improve its ability to generate accurate and properly formatted responses. We opted not to create full MARC records as output, as tools like MarcEdit can easily convert structured text outputs for bibliographic information—with clear field names—into MARC, BIBFRAME, or other metadata schemas. Each prompt was tested across seven bibliographic resources, resulting in a total of 112 outputs.

To compare the performance of each AI model, the following evaluation criteria were applied:

- Prompt responsiveness: How accurately each model responds to various types of prompts, and the extent to which prompt design influences its output.
- Accuracy and consistency: Whether the model provides accurate data, particularly regarding different types of titles, subject headings, and entity authority identifiers.
- Clarification and flexibility: Whether the model asks clarifying questions or suggests alternative outputs.

## FINDINGS AND DISCUSSION

Overall, all tested generative AI models demonstrated an ability to interpret catalogers' needs from the guidance in the prompts. The resources used for the test included both widely cataloged, well-known fiction titles (such as Monograph 1 and Monograph 2) and professional publications with minimal bibliographic records in OCLC (e.g., Serial, Series, and Non-Latin). Across these varied resources and without additional training, the models consistently performed well in identifying and presenting relevant metadata from the images, demonstrating their capacity to follow and retrieve data through both task instructions and visual input. However, their performance differed in three key areas: prompt responsiveness, accuracy and consistency, and clarification and flexibility.

### *Prompt Responsiveness*

All four models generally demonstrated the ability to understand the information requested in the prompts and to generate outputs in the required format. Although each model was tasked with producing the same set of bibliographic elements, the results show that the prompt clearly influenced the performance of AI models, especially when examples were provided. More detailed prompts (such as Prompts 2 and 2e, which include detailed cataloging rules and output examples) tended to elicit more complete, rich, and precise bibliographic outputs.

### *Output Format Consistency*

Although the example used for the prompts explicitly instructed the AI models to produce basic bibliographic information in a structured (non-MARC) human-readable format (see Figure 1), models occasionally returned results in MARC. This inconsistency suggests that while the models recognize and can process MARC, they do not always follow output instructions. When given more

detailed instructions (Prompts 2 and 2e), ChatGPT, DeepSeek, and Copilot consistently followed the specified output format. In contrast, Prompts 1 and 1e led to inconsistent formatting. Gemini was the most unpredictable—returning JSON for series resources and reverting to MARC for non-Latin titles under Prompt 2 (see Figure 2), without asking for clarification.<sup>19</sup> The reasons why JSON was generated were unclear, and no consistent pattern was observed. This unpredictability complicates workflow planning for integrating AI into cataloging.

**Figure 1.** Example included in Prompts 1e and 2e.

Author: Benavente, Jacinto, <https://viaf.org/en/viaf/64015709>

Title: Vidas cruzadas: Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa

Place of Publication: Madrid, Spain

Publisher: Biblioteca Nueva

Date of Publication: 1929

Language: Spanish

Series: Obras de Jacinto Benavente

Series Number: v.2

Content Note: 'Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa.'; 'Estrenada en el Teatro Reina Victoria de Madrid el 30 de marzo de 1929.'

Subject Headings:

- LSCH
  - Spanish drama--20th century, <http://id.loc.gov/authorities/subjects/sh85126242>;
  - Class consciousness, <http://id.loc.gov/authorities/subjects/sh85026683>;
  - Drama, <http://id.loc.gov/authorities/subjects/sh99001612>
- FAST
  - Drama, <http://id.worldcat.org/fast/1423879>
- LCGFT
  - Drama, <http://id.loc.gov/authorities/genreForms/gf2014026297>

**Figure 2.** Gemini-generated response to Prompt 2, formatted in MARC.

```

100 1\${a}Brar, Gursharan
Kaur\${0}http://viaf.org/viaf/26866294/\${e}author.
245 10\${a}Sandarabh sewā = reference service /\${c}Gurasharan Kaur
Brar.
260 \${a}Patiala,\${b}India : \${b}Prita Pablikeśhanaza,\${c}2005.
300 \${a}255 pages ;\${c}22 cm.
505 0\${t}Sandarabha sewā: wiwidh pakha --\${t}Sandarabha sewā dī wiūṁ
kīāṁ --\${t}Tarata sandarabha sewā te dījīṭala kālīna sandarabha sewā
--\${t}Sandarabha prasana utara te jāṁca paṛatāla dā mahatava
--\${t}ṭākūmeṅṭarī surasā mūla --\${t}Manaca mūla --\${t}Sūcanā mūla
--\${t}ḍāiṭarī sīkhana --\${t}Sandarbha --\${t}Mahānakosha te wiśwakosha
--\${t}Dhārasika pustakāṁ te saṁskaraṇa --\${t}Jāidāiṭariāṁ
--\${t}Ākhirakārī sākhī mūla --\${t}Calaṭa hāṁḍabukaṁ ūpara ādhārata
sandarabha mūla, hāṁḍabukaṁ te menūala --\${t}Nītatī mūla
--\${t}ḍūlīṭika mūla --\${t}Wicholiā wiśwakosha mūla --\${t}ḍāiṭarī
wiśwakoshaṁ --\${t}Liparaka wiśwakoshaṁ --\${t}Wiśa wiśwakoshaṁ
--\${t}ḍāiṭarī āḍīo-wīḍīo pustarā --\${t}Calaṭa sandarbha sewā.
546 \${a}In Panjabi.
650 \${0}http://id.loc.gov/authorities/subjects/sh85112114\${a}Reference
services (Libraries)

```

### *Impact of Prompt Design on Output Results*

For our test cases, prompt design impacted the outputs, though not consistently across models. Prompt 2 included explicit RDA rules, such as the instruction to record only the year (YYYY) in the “Date of Publication” field. In two test cases (Monograph 4 and Series) where publication dates included both month and year, all models retained the month when responding to Prompts 1 and 1e, which do not include RDA guidance. With Prompt 2, however, ChatGPT, DeepSeek, and Copilot correctly applied the RDA rule and omitted the month. Gemini failed to do so under Prompt 2 but complied when given Prompt 2e. For Monograph 4 specifically, DeepSeek and ChatGPT correctly provided only the year across all prompts, while Copilot and Gemini included the month in responses to Prompts 1 and 1e. These results suggest that the detailed prompt with examples may provide clearer instructions to generative AI and help reduce deviation from the desired output.

For the non-Latin resource, all models seemed unclear on whether fields like title, author, publisher, place of publication, and content notes should be provided in the original script, translated to English, or transliterated to Latin, with only ChatGPT and Copilot providing any non-Latin script in their response, despite Prompt 2’s clear instructions that “the title proper is transcribed exactly as it appears on the preferred source of information (usually the title page).” This indicates that the prompt must be more detailed about how to transcribe information, particularly for non-Latin script materials.

All prompts asked the AI model to generate subject headings in LCSH. However, three of the four models did not provide subject headings at all in some instances. Only DeepSeek consistently

provided terms labeled as LCSH, though the terms themselves were not always valid LCSH terms. ChatGPT (Serial, Prompt 2), Copilot (Serial, Prompts 2 and 2e), and Gemini (Monograph 1, Prompts 1e and 2) did not provide any LCSH terms. It is unclear why these models failed to generate subject headings for these particular outputs, especially given that they did so for other combinations of prompt and example.

Overall, Gemini and Copilot exhibited notable shortcomings in following and applying prompt instructions. Even with Prompts 2 and 2e, their performance often declined compared to their responses to the simpler Prompt 1. In several cases, these two models failed to distinguish key elements, such as main title versus subtitle, and provided only keywords instead of established subject headings, even though information to populate these elements was clearly present in the scanned input data.<sup>20</sup> They also inconsistently included uniform resource identifiers (URIs) for names and subjects, sometimes omitting them even when they had done so in response to simpler prompts. Additionally, both models struggled to correctly format “Place of Publication” and “Publisher,” producing outputs that failed to align with the prompt instructions and conform to RDA-based guidelines. In contrast, DeepSeek and ChatGPT consistently demonstrated better prompt adherence, showing a clearer understanding of both general instructions and the use of provided examples.

#### *Controlled Vocabularies and Associated URIs*

Compared with Prompts 1 and 1e, Prompts 2 and 2e improved the models’ use of controlled vocabularies. These two prompts led to the inclusion of terms from the Library of Congress Genre/Form Terms (LCGFT), which did not appear at all in results from Prompt 1 (see Table 1). Additionally, Prompts 2 and 2e asked models to supply URIs for entities and subjects using LCSH, Faceted Application of Subject Terminology (FAST), and LCGFT, confirming that well-crafted prompts and examples improve the completeness of metadata output (see Table 2).

**Table 1.** Inclusion of controlled vocabulary terms and keywords generated by different prompts.

AI Model	Prompt	LCSH	FAST	LCGFT	Keywords
ChatGPT	1	27	28	0	41
	1e	20	18	5	36
	2	15	3	1	0
	2e	15	10	7	0
Copilot	1	17	18	0	27
	1e	18	14	1	20
	2	7	1	0	1
	2e	7	5	5	0
DeepSeek	1	26	23	0	37
	1e	22	17	6	21
	2	21	2	10	0
	2e	17	11	8	0
Gemini	1	17	19	0	28
	1e	9	11	0	10
	2	6	0	0	0
	2e	14	16	10	0

**Table 2.** Inclusion of subject URIs generated by different prompts.

AI Model	Prompt	LCSH	FAST	LCGFT	
ChatGPT	1	0	0	0	
	1e	20	18	5	
	2	13	2	1	
	2e	15	10	7	
	Copilot	1	0	0	0
		1e	15	11	1
2		0	0	0	
	2e	7	5	5	
	DeepSeek	1	0	0	0
		1e	22	17	6
2		6	0	1	
	2e	14	9	8	
	Gemini	1	0	0	0
		1e	5	6	0
2		1	0	0	
	2e	12	12	10	

***Accuracy and Consistency***

Accuracy and consistency present a significant challenge for all AI models. While the generated bibliographic information may appear correct at first glance, closer analysis reveals persistent issues with field placement, formatting, and the accuracy and validity of subject headings and URIs.

***Use of Correct Metadata Fields***

All models demonstrated difficulty correctly assigning information to specific descriptive fields as instructed by prompts. For the serial resource, all models except Gemini were able to extract the title value for either the serial as a whole (“Annual Report of the Ontario Historical Society”) or as an individual issue (“Annual Report of the Ontario Historical Society 1910”) from the input images. Gemini failed to recognize the title information for this resource in its response to Prompt 2e. Common issues for all models included placing contributor names, subtitles, copyright statements, and content notes in inappropriate or overly general note fields. In one notable case, Gemini misidentified the publisher as the author, suggesting a misinterpretation of contextual cues in the source image. Among all models, DeepSeek exhibited the most consistent accuracy in assigning data to the correct fields. Copilot also performed reasonably well, though it showed no measurable improvement with more detailed prompts.

***Accuracy and Validity of Subject Headings***

All prompts included a request for LCSH, FAST, and keywords, with Prompts 2 and 2e expanding the field with LCGFT. Given the limited input available to the models, highly nuanced subject analysis was not expected; however, the inclusion of valid, authorized, and reasonably accurate subject headings was considered essential for the subject heading evaluation.

Nevertheless, some headings were not well supported by the original material or would likely not be chosen by professional catalogers. In most cases, the models provided a mix of keywords and subject headings from authority files. Often, the selected keywords focused on only a minor aspect

of the resource, such as a single chapter, or highlighted details that were not the core of the resource. For example, ChatGPT assigned “Ambulance driver” as a keyword in its response to Prompt 1 for Monograph 2, *A Farewell to Arms*. While it is true that the main character’s occupation is an ambulance driver, the novel primarily explores themes of World War I, love, and identity and belonging. Although the keyword is not technically incorrect, it fails to accurately represent the resource as a whole and would not be considered an appropriate subject heading applied through traditional cataloging practices. This shows that cataloging expertise is highly required to select precise, comprehensive, and authoritative subject headings.

The accuracy of the subject headings was inconsistent across AI models. While the models generated subject headings that were semantically appropriate for the resources being described, the headings did not always exist in the authority files to which they were attributed. For example, DeepSeek assigned “Love stories” as an LCSH in its response for Prompt 2, Monograph 2. While the term is contextually appropriate, it is not a valid heading for any of the three prompted vocabularies and therefore should have been included as a keyword instead.

Table 3 below summarizes the total number of subject headings provided by each AI model, along with the percentage of those that were not authorized headings. Our evaluation had to account for compound headings, as each model included them. For FAST and LCGFT, terms could either be valid as authorized headings or not valid, with subdivisions only valid if the FAST heading was fully authorized. For LCSH, we included a third option for partially valid headings, meaning that not all, but at least one, of the subdivision headings was an authorized LCSH. When all subdivision headings were authorized LCSH, these were considered valid. Among the models, Gemini had the highest rate of non-authorized subject headings at 35.29%. This suggests that Gemini is more prone to generating subject headings that are plausible in meaning but are not actually present in established authority files. DeepSeek and ChatGPT outperformed Copilot and Gemini in terms of LCSH quantity and validity.

**Table 3.** Total count of subject headings provided by each model with count of headings authorized in LCSH, FAST, or LCGFT.

AI Model	Total Subject Headings	Non-Authorized Subject Headings (%)
ChatGPT	149	27 (18.12%)
DeepSeek	163	43 (26.38%)
Copilot	93	20 (21.51%)
Gemini	102	36 (35.29%)

#### *Accuracy and Validity of URIs*

While Prompt 1 does not ask the AI models to generate URIs for authorities, the example used in Prompts 1e and 2e and instructions in Prompt 2 included authority URIs. The generated responses to Prompts 1e, 2, and 2e did provide varying URIs for authors and contributors; however, across all models, the majority of these URIs were invalid or incorrect, meaning that either they pointed to an incorrect resource or that no resource existed when attempting to resolve. DeepSeek was the only model to produce one correct VIAF URI for Ernest Hemingway. This may suggest that DeepSeek had this particular URI in its training data, instead of having retrieved or constructed it dynamically, as all other author and contributor URIs were not valid.

In terms of URIs for subject headings, out of 264 URIs included in the output metadata created by all four models, only 18 were completely accurate. Some URIs were accurate for part of a compound heading; for example, “Voyages and travels—Fiction” provided the URI for “Voyages and travels,” but these were not considered valid for our purposes. Sixteen of the correct URIs were provided by DeepSeek, and ChatGPT and Copilot provided one correct URI each. These results suggest that none of the models are trained on the full corpus of LCSH and FAST Linked Data Services, nor can they search these services directly.

### ***Clarification and Flexibility***

Evaluating whether an AI model asks clarifying questions or proposes alternative outputs is critical for understanding its ability to handle ambiguity, avoid incorrect assumptions, and support collaborative, human-in-the-loop workflows. This behavior signals a more advanced grasp of prompt intent and adaptability.

ChatGPT consistently demonstrated this capability. It frequently offered alternative formats and explained why certain data might be missing from its output. DeepSeek also exhibited thoughtful behavior, providing reasoning for the fields it included and explicitly noting omissions. In contrast, Gemini was less inclined to ask clarifying questions or suggest alternatives; instead, it tended to include only minimal explanatory notes. Copilot generated the most concise responses overall, with limited clarifications, usually just asking if a different format was preferred or whether additional fields should be added.

## **CONCLUSION**

This study conducted a small-scale comparative evaluation of four generative AI models: ChatGPT, DeepSeek, Gemini, and Copilot, in generating bibliographic description for diverse resource types. Our evaluation highlights both the promise and the limitations of using current AI tools to support cataloging work. While the sample size in this study is limited and does not support broad statistical generalizations, the observed patterns offer meaningful insights into model behaviors and areas for future exploration.

Across the models, DeepSeek and ChatGPT demonstrated comparatively stronger performance in prompt adherence, accuracy, and validity for generated information and result clarity, particularly when provided with detailed instructions and example outputs. These findings underscore the importance of prompt design: richer, more specific prompts that incorporate cataloging standards, guidance, and examples could lead to improved outputs.

Nonetheless, persistent issues remain. All models struggled with the accuracy and validity of subject headings and URIs, often generating entries that appeared plausible but were ultimately incorrect or non-authorized terms. Misuse of descriptive fields, inconsistencies in applying controlled vocabularies, and failure to reliably process non-Latin scripts suggest that human expertise remains essential to ensure metadata quality and accuracy. Moreover, the variability in model performance underscores the need for human review and intervention in AI-assisted cataloging works.

While generative AI shows some potential as a supplementary tool, it is not capable of replacing catalogers, especially for tasks that require intellectual understanding of the resources and cataloging rules. Metadata creation is not merely a task of transcription or extracting information from a cover page; it requires critical judgment, subject expertise, and contextual understanding that current generative AI models are not capable of replicating. The inability of these tools to

consistently follow detailed cataloging rules, validate authority data, or apply complex subject headings highlights their limitations and immaturity in the library domain. Responsible integration of AI into cataloging workflows must acknowledge these shortcomings and guard against overreliance. Human catalogers remain essential to maintaining the integrity, reliability, and ethical stewardship of bibliographic metadata.

## APPENDIX: PROMPTS

### *Prompt 1*

Your task is to help catalogers create the basic MARC metadata. Please extract information from the provided image: author (last name, first name in order), title, place of publication, publisher, date of publication, language, series (if there is any), series number(if there is any), IF there is a TOC page create a content notes; create subject headings (LCSH and FAST, and Keywords) and any other information you can extract from the page.

### *Prompt 1e*

Your task is to help catalogers create the basic MARC metadata. Please extract information from the provided image: author (last name, first name in order), title, place of publication, publisher, date of publication, language, series (if there is any), series number(if there is any), IF there is a TOC page create a content notes; create subject headings (LCSH and FAST, and Keywords) and any other information you can extract from the page.

Example output looks like below:

- Author: Benavente, Jacinto, <https://viaf.org/en/viaf/64015709>
- Title: Vidas cruzadas: Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa
- Place of Publication: Madrid, Spain
- Publisher: Biblioteca Nueva
- Date of Publication: 1929
- Language: Spanish
- Series: Obras de Jacinto Benavente
- Series Number: v.2
- Content Note: 'Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa.'; 'Estrenada en el Teatro Reina Victoria de Madrid el 30 de marzo de 1929.'
- Subject Headings:
  - LCSH-
    - Spanish drama--20th century, <http://id.loc.gov/authorities/subjects/sh85126242>;
    - Class consciousness, <http://id.loc.gov/authorities/subjects/sh85026683>;
    - Drama, <http://id.loc.gov/authorities/subjects/sh99001612>
  - FAST-
    - Drama, <http://id.worldcat.org/fast/1423879>
  - LCGFT-
    - Drama, <http://id.loc.gov/authorities/genreForms/gf2014026297>

### *Prompt 2*

Your task is to analyze the provided image(s) and extract bibliographic information to create core MARC-style metadata.

Please extract and format the following fields (skip any field that cannot be confidently identified from the image):

1. Author

- Record a creator/author by following guidelines below (RDA 19.2.1.3):
- Format as: Surname, Firstname(s)
- Follow Library of Congress Name Authority rules.
- Please search the [VIAF.org](http://VIAF.org) for matching names. If found, please add the URI of the name as well.

## 2. Title

- Record a title by following guidelines below. RDA 5.1.3
  - Name of a manifestation, that is, a title normally used when naming the manifestation. An alternative title is treated as part of the title proper. A title proper does not include: parallel title proper, other title information, or parallel other title information. A file name or data set name is not considered a title proper unless it is the only title appearing in the manifestation.
  - Source: The title proper is transcribed exactly as it appears on the preferred source of information (usually the title page).
  - Alternative titles (like “or, Something something”) are included as part of the title proper.
  - Parallel titles (in another language/script) and other title information (subtitles, descriptions) are not included in the title proper.
  - No corrections are made to spelling or grammar (e.g., “her first hundred years” is not capitalized unless that’s how it appears).
  - Capitalization and punctuation follow the transcription principle (i.e., reproduce what you see, with minimal normalization).

## 3. Place of Publication

- Record a place of publication associated with the publication, release, or issuing of a resource, include both the local place name (city, town, etc.) and the name of the larger jurisdiction or jurisdictions (state, province, etc., and/or country) if present on the source of information. RDA 2.8.1
  - Record a place of publication in
    - City, State if the place is in USA
    - City, Country if the place is outside of USA
    - If more than one place of publication is named on the source of information, record the place names in the order indicated by the sequence, layout, or typography of the names on the source of information. RDA 2.8.2.4

## 4. Publisher

- Record a name of an agent responsible for publishing, releasing, or issuing a resource. A publisher’s name may be represented by a characterizing word or phrase. RDA 2.8.4.1
  - Record words or phrases indicating the function (other than solely publishing) performed by an agent as they appear on the source of information.
  - If more than one agent is named as a publisher of the manifestation, record the publishers’ names in the order indicated by the sequence, layout, or typography of the names on the source of information. RDA 2.8.4.5

## 5. Date of Publication

- Record a date associated with the publication, release, or issuing of a resource. RDA 2.8.6
  - Record a date of publication in YYYY format. If there is a copyright date available, record the date with Copyrighted YYYY format in addition to the publication date.
  - For multipart monographs and serials, take the beginning and/or ending date of publication from the first and/or last released issue or part, or from another source.
  - For integrating resources, take the beginning and/or ending date of publication from the first and/or last iteration, or from another source.
  - If a date of publication as it appears on the manifestation is not of the Gregorian or Julian calendar, add the corresponding date or dates of the Gregorian or Julian calendar. Indicate that the information was taken from a source outside the manifestation itself.

## 6. Language

- Record a language of expression. RDA 6.11.1.3
  - Use the language of the content in the image (e.g., English, Spanish, French).
  - If a single expression of a work involves more than one language, record each of the languages. RDA 6.11.1.4

## 7. Series (if present)

- Record a title proper of the series. RDA 2.3.1
  - Record the title of the series to which a manifestation belongs. EXAMPLE
    - Bartholomew world travel series
    - Great sacred choruses
    - Allstate simulation film library
    - Sahitya Akademi archives of Indian literature film series
    - Listening, looking, and feeling
  - Exception,
    - If: the manifestation being described consists of two or more issues or parts
    - and numbering that is an integral part of a title proper of series differs from issue to issue or part to part
    - then: omit the numbering from the title proper of series. Use a mark of omission (...) to indicate such an omission. Record the numbering as numbering within the series.
    - EXAMPLE: Publication ... of the Indiana University Research Center in Anthropology, Folklore, and Linguistics

## 8. Series Number (if present)

- If a title proper of series includes numbering as an integral part of the title, transcribe the numbering as part of the title proper of series. RDA 2.3.1
  - Record the numbering of series as appeared on the title page EXAMPLE
    - Publication #122 of the Social Science Education Consortium
    - The twenty-sixth L. Ray Buckendale lecture

## 9. Content Notes

- Record a brief objective summary of the content of the expression and a Table of contents, if there is a table of contents page available.
  - Extract content notes if a Table of Contents is available. List section or chapter titles with an author, if the author information is available. Example:
    - Quark models - Rosner, John.
    - Introduction to gauge theories of the strong, weak, and electromagnetic interactions - Johnson, Kim

## 10. Contributors:

- Record contributors IF those chapters and sections of the table of contents have authors:
  - Format as: Surname, Firstname(s)
  - Follow Library of Congress Name Authority rules.
  - Search the [VIAF.org](https://viaf.org) for matching names. If found, please add the URI of the name as well.

## 11. Subject Headings

- Record a description of the related subject by using either a structured description or an unstructured description (e.g., keywords), as appropriate. RDA 23.4.1.2.3
  - Assign a complement of access points that provide access to at least the primary/essential subject and/or genre/form of the work at the appropriate level of specificity. Assign such access points from an established thesaurus, list, or subject heading system.
  - Use controlled vocabularies from:
    - Library of Congress Subject Headings (LCSH): <https://id.loc.gov/authorities/subjects.html>
    - FAST (Faceted Application of Subject Terminology): <https://fast.oclc.org/searchfast/>
    - Genre/Form Headings: <https://id.loc.gov/authorities/genreForms.html>

### **Prompt 2e**

Your task is to analyze the provided image(s) and extract bibliographic information to create core MARC-style metadata.

Please extract and format the following fields (skip any field that cannot be confidently identified from the image):

#### 1. Author

- Record a creator/author by following guidelines below (RDA 19.2.1.3):
- Format as: Surname, Firstname(s)
- Follow Library of Congress Name Authority rules.
- Please search the [VIAF.org](https://viaf.org) for matching names. If found, please add the URI of the name as well.

#### 2. Title

- Record a title by following guidelines below. RDA 5.1.3

- Name of a manifestation, that is, a title normally used when naming the manifestation. An alternative title is treated as part of the title proper. A title proper does not include: parallel title proper, other title information, or parallel other title information. A file name or data set name is not considered a title proper unless it is the only title appearing in the manifestation.
- Source: The title proper is transcribed exactly as it appears on the preferred source of information (usually the title page).
- Alternative titles (like “or, Something something”) are included as part of the title proper.
- Parallel titles (in another language/script) and other title information (subtitles, descriptions) are not included in the title proper.
- No corrections are made to spelling or grammar (e.g., “her first hundred years” is not capitalized unless that’s how it appears).
- Capitalization and punctuation follow the transcription principle (i.e., reproduce what you see, with minimal normalization).

### 3. Place of Publication

- Record a place of publication associated with the publication, release, or issuing of a resource, include both the local place name (city, town, etc.) and the name of the larger jurisdiction or jurisdictions (state, province, etc., and/or country) if present on the source of information. RDA 2.8.1
  - Record a place of publication in
    - City, State if the place is in USA
    - City, Country if the place is outside of USA
    - If more than one place of publication is named on the source of information, record the place names in the order indicated by the sequence, layout, or typography of the names on the source of information. RDA 2.8.2.4

### 4. Publisher

- Record a name of an agent responsible for publishing, releasing, or issuing a resource. A publisher’s name may be represented by a characterizing word or phrase. RDA 2.8.4.1
  - Record words or phrases indicating the function (other than solely publishing) performed by an agent as they appear on the source of information.
  - If more than one agent is named as a publisher of the manifestation, record the publishers’ names in the order indicated by the sequence, layout, or typography of the names on the source of information. RDA 2.8.4.5

### 5. Date of Publication

- Record a date associated with the publication, release, or issuing of a resource. RDA 2.8.6
  - Record a date of publication in YYYY format. If there is a copyright date available, record the date with Copyrighted YYYY format in addition to the publication date.
  - For multipart monographs and serials, take the beginning and/or ending date of publication from the first and/or last released issue or part, or from another source.
  - For integrating resources, take the beginning and/or ending date of publication from the first and/or last iteration, or from another source.

- If a date of publication as it appears on the manifestation is not of the Gregorian or Julian calendar, add the corresponding date or dates of the Gregorian or Julian calendar. Indicate that the information was taken from a source outside the manifestation itself.

## 6. Language

- Record a language of expression. RDA 6.11.1.3
  - Use the language of the content in the image (e.g., English, Spanish, French).
  - If a single expression of a work involves more than one language, record each of the languages. RDA 6.11.1.4

## 7. Series (if present)

- Record a title proper of the series. RDA 2.3.1
  - Record the title of the series to which a manifestation belongs. EXAMPLE
    - Bartholomew world travel series
    - Great sacred choruses
    - Allstate simulation film library
    - Sahitya Akademi archives of Indian literature film series
    - Listening, looking, and feeling
  - Exception,
    - If: the manifestation being described consists of two or more issues or parts
    - and numbering that is an integral part of a title proper of series differs from issue to issue or part to part
    - then: omit the numbering from the title proper of series. Use a mark of omission (...) to indicate such an omission. Record the numbering as numbering within the series.
    - EXAMPLE: Publication ... of the Indiana University Research Center in Anthropology, Folklore, and Linguistics

## 8. Series Number (if present)

- If a title proper of series includes numbering as an integral part of the title, transcribe the numbering as part of the title proper of series. RDA 2.3.1
  - Record the numbering of series as appeared on the title page EXAMPLE
    - Publication #122 of the Social Science Education Consortium
    - The twenty-sixth L. Ray Buckendale lecture

## 9. Content Notes

- Record a brief objective summary of the content of the expression and a Table of contents, if there is a table of contents page available.
  - Extract content notes if a Table of Contents is available. List section or chapter titles with an author, if the author information is available. Example:
    - Quark models - Rosner, John.
    - Introduction to gauge theories of the strong, weak, and electromagnetic interactions - Johnson, Kim

## 10. Contributors:

- Record contributors IF those chapters and sections of the table of contents have authors:
  - Format as: Surname, Firstname(s)
  - Follow Library of Congress Name Authority rules.
  - Search the [VIAF.org](https://viaf.org) for matching names. If found, please add the URI of the name as well.

## 11. Subject Headings

- Record a description of the related subject by using either a structured description or an unstructured description (e.g., keywords), as appropriate. RDA 23.4.1.2.3
  - Assign a complement of access points that provide access to at least the primary/essential subject and/or genre/form of the work at the appropriate level of specificity. Assign such access points from an established thesaurus, list, or subject heading system.
  - Use controlled vocabularies from:
    - Library of Congress Subject Headings (LCSH): <https://id.loc.gov/authorities/subjects.html>
    - FAST (Faceted Application of Subject Terminology): <https://fast.oclc.org/searchfast/>
    - Genre/Form Headings: <https://id.loc.gov/authorities/genreForms.html>

Example output looks like below:

- Author: Benavente, Jacinto, <https://viaf.org/en/viaf/64015709>
- Title: Vidas cruzadas: Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa
- Place of Publication: Madrid, Spain
- Publisher: Biblioteca Nueva
- Date of Publication: 1929
- Language: Spanish
- Series: Obras de Jacinto Benavente
- Series Number: v.2
- Content Note: 'Cinedrama en dos partes, dividida la primera en diez cuadros y la segunda en tres y un epílogo, y en prosa.'; 'Estrenada en el Teatro Reina Victoria de Madrid el 30 de marzo de 1929.'
- Subject Headings:
  - LCSH-
    - Spanish drama--20th century, <http://id.loc.gov/authorities/subjects/sh85126242>;
    - Class consciousness, <http://id.loc.gov/authorities/subjects/sh85026683>;
    - Drama, <http://id.loc.gov/authorities/subjects/sh99001612>
  - FAST-
    - Drama, <http://id.worldcat.org/fast/1423879>
  - LCGFT-
    - Drama, <http://id.loc.gov/authorities/genreForms/gf2014026297>

---

**ENDNOTES**

- <sup>1</sup> Osma Suominen, "Annif: DIY Automated Subject Indexing Using Multiple Algorithms," *LIBER Quarterly: The Journal of the Association of European Research Libraries* 29, no. 1 (2019): 1–25, <https://doi.org/10.18352/lq.10285>.
- <sup>2</sup> Koraljka Golub, "Automated Subject Indexing: An Overview," *Cataloging & Classification Quarterly* 59, no. 8 (2021): 702–19, <https://doi.org/10.1080/01639374.2021.2012311>.
- <sup>3</sup> Osma Suominen, Juho Inkinen, and Mona Lehtinen, "Annif and Finto AI: Developing and Implementing Automated Subject Indexing," *Bibliographic Control in the Digital Ecosystem, JLIS.It* 13, no. 1 (2022): 265–82, <https://doi.org/10.4403/jlis.it-12740>.
- <sup>4</sup> Koraljka Golub et al., "Automated Dewey Decimal Classification of Swedish Library Metadata Using Annif Software," *Journal of Documentation* 80, no. 5 (2024): 1057–79, <https://doi.org/10.1108/JD-01-2022-0026>.
- <sup>5</sup> Charlene Chou and Tony Chu, "An Analysis of BERT (NLP) for Assisted Subject Indexing for Project Gutenberg," *Cataloging & Classification Quarterly* 60, no. 8 (2022): 807–35, <https://doi.org/10.1080/01639374.2022.2138666>.
- <sup>6</sup> Jenny Bodenhamer, *Reliability and Usability of ChatGPT for Library Metadata* (Oklahoma State University, 2023), <https://hdl.handle.net/20.500.14446/339626>.
- <sup>7</sup> Shoichi Taniguchi, "Creating and Evaluating MARC 21 Bibliographic Records Using ChatGPT," *Cataloging & Classification Quarterly* 62, no. 5 (2024): 527–46, <https://doi.org/10.1080/01639374.2024.2394513>.
- <sup>8</sup> Eric H. C. Chow, T. J. Kao, and Xiaoli Li, "An Experiment with the Use of ChatGPT for LCSH Subject Assignment on Electronic Theses and Dissertations," *Cataloging & Classification Quarterly* 62, no. 5 (2024): 574–88, <https://doi.org/10.1080/01639374.2024.2394516>.
- <sup>9</sup> Brian Dobreski and Christopher Hastings, "AI Chatbots and Subject Cataloging: A Performance Test," *Library Resources & Technical Services* 69, no. 2 (2025), <https://doi.org/10.5860/lrts.69n2.8440>.
- <sup>10</sup> Richard Brzustowicz, "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging," *Information Technology and Libraries* 42, no. 3 (2023), <https://doi.org/10.5860/ital.v42i3.16295>.
- <sup>11</sup> Brzustowicz, "From ChatGPT to CatGPT."
- <sup>12</sup> Brzustowicz, "From ChatGPT to CatGPT."
- <sup>13</sup> Christine DeZelar-Tiedman, "Response to 'From ChatGPT to CatGPT,'" *Information Technology and Libraries* 42, no. 4 (2023), <https://doi.org/10.5860/ital.v42i4.16991>; Tess Amram, Robin Goodfellow Malamud, and Cheryl Hollingsworth, "Response to 'From ChatGPT to CatGPT,'" *Information Technology and Libraries* 42, no. 4 (2023), <https://doi.org/10.5860/ital.v42i4.16983>; David Floyd, "Response to 'From ChatGPT to CatGPT,'" *Information Technology and Libraries* 42, no. 4 (2023), <https://doi.org/10.5860/ital.v42i4.16995>.
- <sup>14</sup> Heather Moulaison-Sandy and Zach Coble, "Leveraging AI in Cataloging: What Works, and Why?," *Technical Services Quarterly* 41, no. 4 (2024): 375–83, <https://doi.org/10.1080/07317131.2024.2394912>.
- <sup>15</sup> Vyacheslav Zavalin and Oksana L. Zavalina, "Are We There Yet? Evaluation of AI-Generated Metadata for Online Information Resources," *Information Research an International Electronic Journal* 30 (iConf): 732–40, <https://doi.org/10.47989/ir30iConf47215>.
- <sup>16</sup> Lili Daie, Sungmin Park, Mary Beth Weber, and Alicia Zalusky, "Beta-Testing the Alma AI Metadata Assistant: Early Access Insights from Libraries and Ex Libris," presented at the ELUNA 2025 Conference, Atlanta, GA, June 20, 2025, <https://elunaconf2025.dryfta.com/program-schedule/program-detail/51/beta-testing-the-alma-ai-metadata-assistant-early-access-insights-from-libraries-and-ex-libris>.
- <sup>17</sup> The scanned images for the Panjabi book include the romanized title, author, imprint, and language information: <https://github.com/gretaheng/Generative-AI-Meets-Cataloging-Practice/tree/main/Input%20data/Non-Latin-Punjabi>.

- <sup>18</sup> The scanned input data can be found in the project's GitHub repository: <https://github.com/gretaheng/Generative-AI-Meets-Cataloging-Practice/tree/main/Input%20data>.
- <sup>19</sup> The generated JSON record example can be viewed at <https://github.com/gretaheng/Generative-AI-Meets-Cataloging-Practice/blob/main/AI%20Generated%20Results/English-Serial/Eng-Serial-Gemeni-Prompt-2.docx>.
- <sup>20</sup> For example, although the subtitle is clearly provided in the scanned PDF (<https://github.com/gretaheng/Generative-AI-Meets-Cataloging-Practice/blob/main/Input%20data/English-Mono3-NonFiction/Book%204-Cover.pdf>), Gemini's output for Prompt 1 captured only the main title and omitted the subtitle (see <https://github.com/gretaheng/Generative-AI-Meets-Cataloging-Practice/blob/main/AI%20Generated%20Results/English-Mono3-NonFiction/English-Mono3-NonFiction-Gemeni-Prompt-1.docx>).