Journal of the University Librarians Association of Sri Lanka, Vol. 27, Issue 2, July 2024, 227–256, DOI: https://doi.org/10.4038/jula.v27i2.8083

Using Generative AI for Bibliographic Description: A Study with ChatGPT 4

R.C.G. Gamage[1] and Priyanwada Wanigasooriya[2]

Abstract

This study explores the use of Generative Artificial Intelligence (GAI) in bibliographic description in university library catalogs. It focuses on GAI's potential in enhancing efficiency and maintaining consistency of bibliographic description while complying with cataloging standards such as MARC21, AACR2 and RDA. The study followed a qualitative methodology, examining 10 use cases in metadata extraction, RDA compliance of GAI generated catalog records, and error checking of existing catalog records. The study used OpenAI's ChatGPT 4 to execute the practical tests. The study concludes that GAI can significantly assist in improving cataloging efficiency. It was highly satisfactory in error correction of existing MARC21 records. GAI was also a prospective tool for educating novice librarians on the use of MARC21 and RDA standards. CATMELK, a custom GPT model created with ChatGPT 4 could effectively convert images from books' title, copyright and cover pages into RDA compliant MARC21 records. However, when the tests were conducted in January 2024, ChatGPT 4 faced difficulties in direct conversion of Sinhala and Tamil data from images to MARC2 1 records. Considering the rapid development of GAI technology and applications, it is recommended to repeat the tests on a continuing basis.

Keywords: AACR2, MARC21, RDA, Artificial Intelligence, ChatGPT

---

[1] Senior Lecturer, National Institute of Library and Information Sciences, University of Colombo.

Email: ruwan@nilis.cmb.ac.lk https://orcid.org/0000-0002-7081-8794

[1] Senior Lecturer, Department of Library and Information Science, University of Kelaniya.

Email: priyanwada@kln.ac.lk https://orcid.org/0000-0001-9219-5025

Received: 30 May 2024, Accepted revised version: 25 July 2024

Introduction

Library catalogs are not merely listings of library holdings; they are crucial for research, discovery, and learning. Maintaining a discoverable catalog requires bibliographic control and bibliographic description. Bibliographic control refers to the system used in the creation and maintenance of a library's bibliographic records (Baldwin, 1976). It involves maintaining consistency of record keeping, which helps achieve control of heading usage in catalogs. Bibliographic description is adding controlled metadata to represent a library material in the catalog. The aim of this study is to explore the capability of Generative AI (GAI) for efficient creation and management of catalog records.

Bibliographic description has traditionally been a balance between rule-based actions and expert decision-making. As such, a catalog record is a mix of structured data, based on established rules and subjective data shaped by the catalogers' expertise and experience. Therefore, adherence to bibliographic description standards is an essential element in improving discoverability (Cabonero et al., 2021; Chaplin, 1956; Fell & Ou, 2022; Neatrour & Hebron, 2023; Thompson & Hoover, 2023; Woolcott et al., 2021).

Standards such as the Anglo-American Cataloging Rules (AACR2) (Joint Steering Committee for Revision of AACR & American Library Association, 2002), and Resource Description and Access (RDA) (Joint Steering Committee for Development of RDA et al., 2015) guide the cataloging processes. The MARC21 format (*MARC 21 Format for Bibliographic Data*, 2023), a format of electronic data exchange in libraries, enables catalog records to be shared and understood not only by catalogers but also by computers, thus facilitating seamless information exchange. The National Library of Sri Lanka (NLSL) introduced the Descriptive Bibliographic

Framework (DBIB) as a guideline for selecting essential MARC21 tags in bibliographic description of various formats such as books and monographs, serial publications, and theses and dissertations, within the Sri Lankan context (Gamage et al., 2020, 2024).

Research Problem

Catalogers face several challenges in bibliographic description including inconsistency of data sources (Visser et al., 2021), lack of competency in foreign languages (Bailey & Shou, 2021; Olson, 2014), complexity of cataloging rules (Perry, 2023), technological limitations of existing systems (Seikel & Steele, 2020), the volume of work (Perry, 2023), subjectivity in cataloging (Perry, 2023), data entry errors (Kushwah & Singh, 2022), maintaining updated records (Antelman et al., 2006), standardization across systems (Hines, 2021), and the need for training and professional development (Tosaka & Park, 2018). Breeding (2022) notes that the time required by a library cataloger to create a MARC record varies from one hour to 20 minutes.

The above issues that catalogers face in bibliographic description can be broadly categorized into three areas: extraction of bibliographic details, conversion of them into MARC records, and management of records.

Research confirms that these have affected the quality of catalog records in Sri Lankan university libraries as well. A study which reviewed cataloging practices in four long established Sri Lankan university libraries highlighted a critical need for the standardization of cataloging practices aligned with international best practices (Wanigasooriya & Gamage, 2024). The common errors identified in this study were incorrect use of language Code (Tag 041), incorrect main headings and added entries (Tags 100, 700),

incorrect use of responsibility area and its author name format (Tag 245), structural errors in imprint data (Tag 260), and the use of uncontrolled vocabulary in subject headings (Tag 650). The study suggests a gap in adhering to AACR2, and MARC21 standards by the tested Sri Lankan University library online public access catalogs (OPAC). Furthermore, the study revealed that none of these catalogs were compliant with RDA. Sampath (2011) and Ketheeswaren (2017) also indicated the gaps in data entering into Sri Lankan University catalogs.

herefore, an intervention is required to make bibliographic description efficient, and effective. Thus, the aim of the research is to explore the ability of GAI to achieve this purpose.

Research Objectives and Questions

The research objectives and questions were derived from the above identified research problem. Also, using GAI for extracting metadata involves use of OCR technology. The vernacular languages Sinhala and Tamil are low resourced when compared to English (Vasantharajan et al., 2021), leading to low accuracy of capturing. Therefore, a fourth objective was added to explore the ability of GAI to convert Sinhala and Tamil books.

Table 1

Research objectives and research questions of the study

Aspect	**Research Objectives**	**Research Questions**
1. Data Extraction	1. To observe the capability of GAI in extracting bibliographic	1.1 Can GAI extract bibliographic details from the Web by ISBN?

	details from various sources.	1.2 Can GAI extract bibliographic details from a database by ISBN?
2. MARC Record Conversion	2. To evaluate the effectiveness of GAI in converting bibliographic details to MARC records, including those for foreign language materials.	2.1 How effectively does GAI convert book cover and/or title page details to a MARC record?
		2.2 How effectively does GAI convert book cover and/or title page details to MARC record in other foreign language materials?
		2.3 How effectively does GAI convert a catalog card to a MARC record?
		2.4 How effectively does GAI convert a National Bibliography page to MARC record?
		2.5 How effectively does GAI create DBIB compliant MARC21 records?
		2.6 How effectively does GAI create RDA compliant MARC21 records with a custom built GPT?
	3. To investigate the ability of GAI to manage	3.1 Can GAI display MARC records from a file?

3. MARC Record Management	and correct MARC records.	3.2 How proficiently does GAI correct errors in existing MARC records?
4. Capturing vernacular languages	4. To observe the ability of GAI to capture Sinhala and Tamil	4.1 Can GAI capture Sinhala and Tamil books, and convert to MARC records?

Generative AI Applications

Generative AI (GAI) is a type of artificial intelligence capable of creating new content, such as text, images, or music, by learning patterns from existing data and generating new outputs. A prominent example of GAI is GPT, which stands for "Generative Pre-trained Transformer." GPT is designed to understand a given input reasonably mirroring a human brain, and produce relevant output.

GAI chat applications, such as AI chatbots, facilitate real-time conversations and respond to user queries referred to as 'prompts'.

The chosen GAI tool for this study was OpenAI's ChatGPT 4. The reason was its rank as one of the best AI chatbots by December 2023 (*Top 5 AI Chatbots That You Need to Try*, 2023).

In the prompt box where conversations are input, ChatGPT 4 has a ‘paperclip icon’ with file attachment capability (Figure 1).

Figure 1

ChatGPT Version 4 Data Input field.

Researchers have extensively discussed the predictive power of GAI (Barbaglia et al., 2021; Management Solutions, 2018; Zihni et al., 2022). Therefore, authors observed that GAI has the potential to integrate with library cataloging, as a measure to improve cataloging efficiency and consistency. This study aims to explore the potential roles and benefits of GAI in enhancing cataloging methods in university libraries, with a focus on book data.

In university libraries, the volume of resources and the need for accurate cataloging is essential. Therefore, GAI's capability to standardize entries and accelerate technical services is particularly of interest. The research problem focuses on investigating whether GAI can effectively address these challenges, thus contributing to the improvement of library service quality and accessibility.

Review of Literature

AI has been applied in various fields since its inception in 1950s. Its applications range from healthcare (Varghese et al., 2023) to human resource management (Bhatt & Shah, 2023; Mer, 2023; Singh & Chouhan, 2023), marketing (Malhotra & Ramalingam, 2023; Paschen et al., 2019), and educational performance assessment (Halagatti et al., 2023). However, the application of AI in the library and information sector is still in a development stage, with research gradually increasing in this area.

The rise and development of GAI offers a promising solution to the challenges faced by catalogers. For example, Metz (2024) visualized the goal of building systems that methodically solve problems, mirroring human reasoning through a series of logical steps. Lowagie's (2023) studies on creating automated processes that incorporate AI capabilities signifies a substantial advancement in improving cataloging efficiency and accuracy

while adhering to metadata standards. Going beyond the cataloging practice level to the standardizing level, Gamage and Wanigasooriya (2023) demonstrated the high similarity of a machine generated DBIB framework with the human created framework recommended by the NLSL.

Xu, Feng and Chen (2023) used a randomized online experiment to compare searching capability of GAI. They used ChatGPT-like and Google Search-like tools in the study. It revealed that the 'ChatGPT' group spent less time on tasks with no significant difference in overall performance. However, it was less effective in fact-checking tasks. Abba (2024) states that some African university libraries are using ChatGPT 4 as a knowledge base for 'cataloguing information'. However, ChatGPT 4 cannot directly query a database (Hook, 2023). Instead it needs GPT4 API to do the task (*Langchain Tutorial. Query a Database with OpenAI's ChatGPT*, 2023).

Brzustowicz (2023) could create RDA Compliant MARC21 records using GAI, but his experiments have been criticized by other researchers for the lack of rigor (Amram et al., 2023; Floyd, 2023). ChatGPT has also demonstrated superior language translating ability when compared with a purpose built application (Ghassemiazghandi, 2024).

Bandi et al. (2023) states that image-to-text models are advancing enabling not only to convert images to text, but also the reverse mechanism as demonstrated by DALL-E. These have combined the power of computer vision, natural language processing, and attention mechanisms.

Researchers have confirmed the utility of ChatGPT's custom GTP models too (Almasre, 2024; Zhao et al., 2024).

Methodology

Studying different approaches of standardizing and enhancing the efficiency and effectiveness of bibliographic description is explorative in nature. Our focus was on the process over metrics, and flexibility and adaptability over strict protocols in the research design. A contextual understanding was necessary for identifying practical solutions and improvements. Therefore, a task-based approach was employed to answer the research questions. Using task-based tests, the study could simulate real-world cataloging scenarios and gather detailed observations on how various tools and methods perform in practice. This inevitably resulted in following the qualitative research methodology.

Since our focus was on working on multiple files to save time on cataloging functions (eg: extracting bibliographic records from multiple sources), we selected ChatGPT 4 as the GAI tool.

All tests were conducted in January 2024. By that time, ChatGPT 4 had the capacity to search the Web for additional information. In addition, it allowed creation of custom GPTs which the creator could build and configure. Therefore, a custom built GPT called CATMELK was created. It is publicly available at https://chatgpt.com/g/g-YQm6gByUe-catmelk. CATMELK can create RDA-compliant catalog records from images captured through a mobile phone, without the need of human mediated textual prompts.

Specific tasks and scenarios were designed to test the required functionality in relation to bibliographic description. Accordingly, the study utilized 49 prompts to carry out 10 tasks to observe reaction to the 4 aspects of activity (Tables 2 and 3). Tasks 1-2 focused on extraction of bibliographic details, 3-8 on conversion into MARC records, and 9-10 on records

management. Task 8 was completed using CATMELK. Since foreign language materials are also required to be added to collections, some materials in common foreign languages also were included in Task 4.

In some of the tasks, multiple inputs were included at once (for example, five ISBN numbers in a single prompt, or five cover page image files as attachments), and in some of the tasks a single input was added. These have been clearly mentioned under the tasks.

By exploring these use cases of GAI in reducing errors and streamlining cataloging processes, the study aimed to provide insights into the feasibility and effectiveness of integrating GAI technologies in library cataloging practice.

It should be noted that the tests were representative but not exhaustive, hence the selected research methodology. It ensured depth and contextual understanding, thereby maintaining research rigor.

Table 2

Mapping tasks and tests to the research questions

Research Questions	**Tasks**	**Task Description**
1.1, 4.1	TASK1	Extract bibliographic details from the Web by ISBN
1.2, 4.1	TASK2	Extract bibliographic details from the WorldCat database by Title and ISBN
2.1, 4.1	TASK3	Convert book cover and/or title page details to MARC record

2.2, 4.1	TASK4	Convert book cover and/or title page details to MARC record in other foreign language materials
2.3, 4.1	TASK5	Convert catalog card to MARC record
2.4, 4.1	TASK6	Convert National Bibliography page to MARC record
2.5	TASK7	Create DBIB compliant MARC21 records
2.6, 4.1	TASK8	Create RDA compliant MARC21 records with CATMELK
3.1	TASK9	Display MARC records from file
3.2	TASK10	Existing MARC record error correction

In the primary data collection phase, researchers directly interacted with GAI applications to see the initial response of AI applications. Subsequently, every test, results, and observations were carefully documented, ensuring comprehensive record-keeping. That followed with a thematic and content analysis, following qualitative analysis methodology. Afterwards, researchers identified patterns, themes, and insights, thereby facilitating meaningful conclusions about the applications' performance and user experience. Lastly, the researchers reflected on their roles and acknowledged any biases or assumptions that might have influenced the findings.

Results

We assessed the performance of ChatGPT 4 in self-performing the above designated tasks. Each task included several prompts fed into ChatGPT 4 (Table 3). Sharable URLs of the complete conversation of each session have been given below the task, whenever available. By the time of the study, ChatGPT did not support the sharing of sessions that include image uploads.

Thus the prompts of tasks which include image files have been given as complete as possible.

Table 3

Overview of Tests Conducted on ChatGPT 4.

Task	**Test Description**	**Process**		
		Sam	***Prom***	***Res***
TASK1	Extract bibliographic details from the Web by ISBN (OpenAI, 2024a)	10	3	*
	English Multiple/Single			√
	Sinhala and Tamil			X
TASK2	Extract bibliographic details from the WorldCat database by Title and ISBN (OpenAI, 2024b)	5	4	X
TASK3	Convert book cover and/or title page details to MARC record (OpenAI, 2024c)	11	3	*
	English Multiple			X
	English Single			√
	Sinhala and Tamil			*
TASK4	Convert book cover and/or title page details to MARC record in other foreign language materials (OpenAI, 2024d)	5	5	√
TASK5	Convert catalog card images to MARC records (OpenAI, 2024e)	10	4	*
	English Multiple/Single			√

	Sinhala and Tamil			X
TASK6	Convert National Bibliography page to MARC record (OpenAI, 2024f)	1	3	*
	English multiple/single			√
	Sinhala and Tamil			X
TASK7	Create DBIB compliant MARC21 records (OpenAI, 2024g)	1	7	√
TASK8	Create RDA Compliant MARC21 records with CATMELK (OpenAI, 2024h)	2	2	*
	English book image to RDA compliant record			√
	Sinhala and Tamil			*
TASK9	Display MARC records from file (OpenAI, 2024i)	4	5	*
	.mrc, .mrk, and .txt			√
	.iso			X
TASK10	Existing MARC record error correction (OpenAI, 2024j, 2024k)	10	13	√
	Punctuation marks identified			√
	Use of Language Code (Tag 041)			√
	Main headings and added entries (100, 700)			√
	Responsibility area and its author name format (245)			√
	Structural errors in imprint data (260)			√

	Controlled vocabulary in subject headings (650)	√
	Correct errors in multiple records	√

Note: The terms 'multiple' and 'single' refer to inputting several records simultaneously or individually. Sam, Prom, and Res are the number of samples tested, number of prompts used, and the result, respectively. Outcomes were indicated as success (√), failure (X), or mixed results (*).

Remarkably, the study highlighted the capability of these applications to streamline cataloging processes, and enhance compliance with MARC21 and RDA standards. Positive results were evident in the effectiveness of Optical Character Recognition (OCR) and web-based metadata retrieval too. However, challenges were also noted in performing such actions specially with the OCR capability for local language texts, unconventional presentation of bibliographic descriptions in Chinese materials, and inability of directly accessing external databases. GAI could overcome the local language conversion issue to some extent. A detailed description of the test results follow.

Extract Bibliographic Details from the Web by ISBN (TASK1)

A common method of searching for bibliographic information is by searching for the International Standard Book Number (ISBN). ChatGPT 4 successfully extracted metadata for ISBNs from the WWW. However, we noted challenges in extracting accurate information for certain ISBNs, especially for books in local languages and ones published within Sri Lanka. For example, it couldn't identify details of the Sinhala title with ISBN 9556460039 and the local English title with ISBN 8170031486.

Extract Bibliographic Details from the WorldCat Database by Title and ISBN (TASK2)

Worldcat.org is the online union catalog of materials cataloged by OCLC member libraries and institutions. It is a publicly searchable database. We faced limitations here, as direct access to WorldCat was not possible for GAI Applications.

Convert Book Details to MARC Record (TASK3)

Information crucial for cataloging is often found in specific sections of library materials, particularly on their initial pages. The title page, for instance, provides essential details such as the Title, Authors, and Imprint. On the verso of title page, copyright and edition-related information is presented, and in some cases, Cataloging in Publication (CIP) data also might be available. The successful completion of this task was evident, especially when structured information like CIP data is available. The system showcased its proficiency in OCR technology, exhibiting effectiveness in handling English records. However, OCR for Sinhala records almost always had difficulties except in a few instances. Often the following message was received; “It seems there was an error in processing the images due to a problem with the OCR language setting. The language 'sin' for Sinhala may not be supported in the current OCR setup.” OCR for Tamil brought mixed results.

Convert Book Cover and/or Title Page Details to MARC Record in Foreign Language Materials (TASK4)

Copyright pages of publications in Arabic, Chinese, French, Korean and Spanish were fed, asking to create MARC21 records. Record creation was successful. ChatGPT was able to recognize Chinese characters, but it was

challenging due to top to bottom writing in some cases, and bibliographic details not being available on a single page etc.

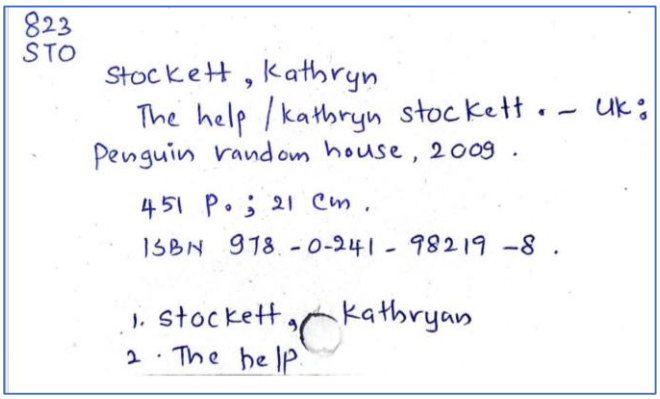
Convert Catalog Card to MARC Record (TASK5)

Some libraries have unconverted traditional catalog cards, of which collections are hidden until they are included in an online catalog. Efficient conversion of the information on these cards is necessary. ChatGPT 4 was successful in converting catalog card data into RDA compliant MARC21 records in a single command. However, the success was dependent on the clarity of the record and OCR capabilities (Figures 3 and 4). After attaching the image file, authors prompted with "Using the attached library catalog card, prepare an RDA compliant MARC21 record". The ChatGPT 4-generated MARC21 record complied with the standards, and logical decision making was also evident. It had correctly omitted irrelevant information in the card. For example, manual catalogers often use the same card format for all tracings, including Title and Author added entries, as shown in Figure 2. The RDA-compliant record as shown in Figure 3 has omitted such added entries since enough access points are already available.

It was also observed that ChatGPT 4 voluntarily accumulate additional bibliographic information from the Web. For example, the generated record populated values for 520 and 650 MARC tags.

Figure 2

Image of a hand-written catalog card captured from a mobile phone camera, used for metadata extraction..



(Source: Handwritten by authors)

Figure 3

MARC record created by ChatGPT4 from the catalog card in Figure 2.

=LDR 00000nam 2200000 4500
=001 123456789
=005 20240508123456.0
=008 240508s2009 enk||||| |||| | eng|d
=020 \\$a 9780241982198
=040 \\$a [Library symbol] $b eng $e rda $c [Library symbol]
=100 1\$a Stockett, Kathryn, $e author.
=245 14$a The Help / $c Kathryn Stockett.
=264 \1$a London : $b Penguin Random House, $c 2009.
=300 \\$a 451 pages ; $c 21 cm.
=336 \\$a text $b txt $2 rdacontent
=337 \\$a unmediated $b n $2 rdamedia
=338 \\$a volume $b nc $2 rdacarrier
=520 \\$a A poignant and powerful story set in the early 1960s in Jackson, Mississippi, focusing on the civil rights movement and the interactions between African American maids and their white employers.
=650 \0$a African American women $x Employment $z Mississippi $z Jackson $v Fiction.
=650 \0$a Civil rights movements $z United States $v Fiction.
=655 \7$a Historical fiction. $2 lcgft

An attempt to create MARC21 records from Sinhala medium catalog cards was unsuccessful, citing inability to interpret Sinhala script.

Convert National Bibliography Page to MARC Records (TASK6)

National Bibliographies are often considered authoritative sources for cataloging. Direct conversion of these to MARC format is helpful in some cases. English language pages from the Sri Lanka National Bibliography (SLNB) were successfully converted to a list of records. However, challenges were encountered in accurately extracting metadata from Sinhala and Tamil text due to OCR limitations and the heavy use of abbreviations in the SLNB.

Create MARC21 Records Following Descriptive Bibliographic Framework (DBIB) Guidelines (TASK7)

ChatGPT 4 created MARC21 records for books and monographs aligned with the DBIB framework as recommended by the National Library of Sri Lanka, when instructions on the essential tags were provided in the prompt. Instructions didn't have to be repeated in a single session. Once provided, it adhered to the instructions on the prompts that followed in the same session.

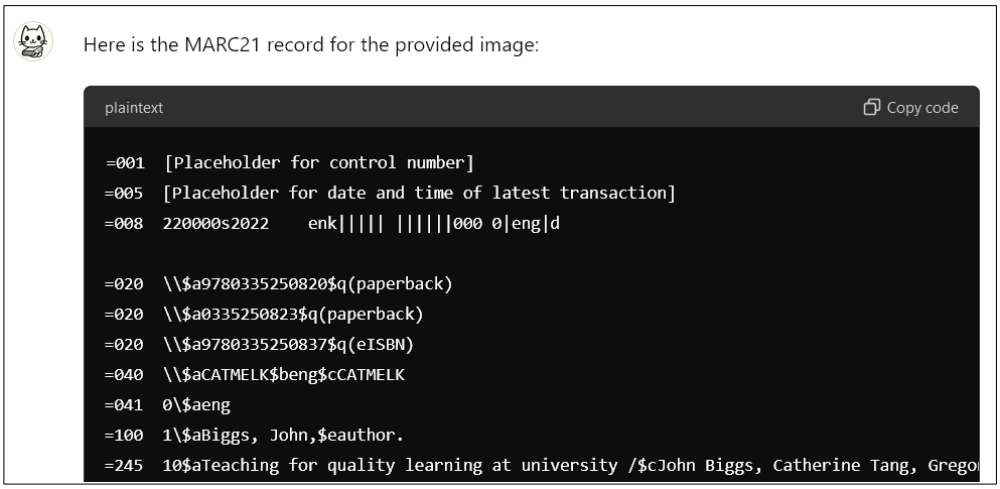
Create RDA Compliant MARC21 Records from Custom Built GPT (TASK8)

CATMELK was configured to create RDA and DBIB compliant records for book image prompts. When the details were comprehensive (for example, from a combination of the cover page, title page, and copyright page), it created nearly complete records. However, the catalogers and the library management system must provide additional details for which the GPT Model leaves placeholders (Figure 4). In addition, the model provides notes

with explanations for the decisions taken, making it helpful for novice catalogers to learn and understand the reasoning behind each description.

Figure 4

Screen shot from CATMELK GPT Model, displaying a section of the code



Display MARC Records from Files (TASK9)

MARC records available in some online catalogs can be downloaded to a PC for error correction and uploading back to the OPAC. MARC records can be downloaded as binary (eg:.mrc and .iso) and non-binary (.mrk, and .txt) files. Opening non-binary files and displaying data was straightforward. However, opening binary files was restrictive. ChatGPT 4 could open human readable parts of a .mrc file, but it was unable to encode an .iso file stating its complexity. These binary files can be easily opened on specialized applications such as MarcEdit and converted to non-binary formats.

Existing MARC Record Error Correction (TASK10)

Even without downloading files, ChatGPT 4 could effectively check records for errors, achieving higher success rates when handling tasks sequentially, such as performing OCR followed by MARC record creation

within the same conversation. The applications could accurately identify and correct errors in 10 consecutive MARC records presented in mnemonic (normal view) format.

Discussion, Conclusion and Recommendations

This study explored the application of Generative AI, particularly ChatGPT 4, for enhancing bibliographic description in university libraries in Sri Lanka. Its focus was on cataloging books. The examination included various tasks carried out under four main aspects. These were data extraction, MARC record conversion & compliance with standards, MARC records management, and working with vernacular languages. The results showed varying degrees of success, highlighting both the potential and limitations of GAI in enhancing bibliographic description function in university library online catalogs. Tasks 7-10 can be further automated to complete functions from capturing new information and error correction of existing records, to importing to the OPAC.

Data Extraction

The tasks involved extracting bibliographic details from the Web and the WorldCat database. The results were mixed, mostly negative with local languages. However, the results aligned with the observations by Xu, Feng, & Chen (2023), who showed that while GAI, specifically ChatGPT, can effectively reduce the time spent on search tasks, its fact-checking capabilities are limited. ChatGPT 4 performed well in extracting details in English but struggled with non-English tasks. As stated by Hook (2023), challenges persist in querying databases directly without additional API support. However, creating records for local languages faced challenges. These may evolve with

the new models of GAI such as ChatGPT-4o, which was far more successful in handling local languages in the recently conducted initial tests.

MARC Record Conversion

The conversion of book cover, title page, copyright page, catalog card, and National Bibliography page images into MARC records also showed mixed results. Just like Lowagie (2023) and Bandi et al. (2023) noted, AI capabilities for converting images to text were evident. Yet our study found that the effectiveness varies by language and task complexity. Single input instances in English were generally successful, but multiple input instances and non-English tasks had lower success rates. This suggests that step by step instructions to ChatGPT with single input task per prompt during a particular conversation can yield more results in bibliographic description, rather than asking multiple tasks to complete within a single prompt.

MARC21 and RDA Compliance

The creation of MARC21 records following specific guidelines showed promising results, especially for English records. This aligns with Brzustowicz (2023) on the ability of GAI in creating MARC21 and RDA compliant records. As Metz (2024) highlighted, human-like reasoning was evident in making decisions, for example, omitting non-RDA compliant data in the result from among the data collected from a traditional catalog card. As demonstrated by CATMELK and confirmed by Almasre (2024) and Zhao et al. (2024), custom GTPs have a special place in the GAI landscape since they can be pre-tuned for specific tasks excluding the need for prompt writing, making the task more efficient.

MARC Record Management

For MARC record management, the tasks included displaying and correcting existing records. The display of MARC records from files was mostly successful, except for .iso files. Error correction tasks were effective in identifying punctuation marks, structural errors, and controlled vocabulary. Handling responsibility areas, and language codes was also successful.

Conclusion, Recommendations, and Acknowledgement of Biases and Limitations

The study concludes that GAI has a significant potential as an assistant in library cataloging, particularly in saving time for routine cataloging tasks and serving as an educational tool for standards like RDA. It is advantageous for novice catalogers, facilitating their learning and adherence to cataloging standards, thereby improving the efficiency of cataloging. It was evident by the fact that CATMELK GPT Model provided descriptive notes on its actions. The study also highlights the transformative impact of GAI on cataloging efficiency and the need for ongoing exploration and refinement in integrating AI into library workflows.

A notable bias existed in the local language processing capability testing, with a preference for Sinhala content over Tamil due to the authors' lack of proficiency in Tamil. This limitation means the results cannot be generalized to Tamil medium contexts. It should also be noted that Sinhala is a language written and spoken by a population of around 13 million (Abeywickrama & Mendis, 2013) mainly within Sri Lanka while Tamil is written and spoken by more than 64 million in the world (Encyclopedia Britannica, Inc., 2024). Therefore, there is a disparity in language availability on the Internet.

Furthermore, the premium version of ChatGPT was available to the researchers, making it the preferred medium for testing. In January 2024 when the tests were conducted, ChatGPT 4 demonstrated significant capabilities in natural language processing tasks, such as text generation, translation, and basic data extraction. By May 2024, advancements in ChatGPT's development have markedly improved its performance. When asked to mark differences between January and May 2024 versions, ChatGPT 4o, the most recent version of the application highlighted enhanced multilingual support, better integration with real-time databases, improved verification algorithms and image-to-text models, and advances in OCR technology (OpenAI, 2024).

Therefore, it is important to note that the field of Generative AI is rapidly evolving. The initial biases, the subsequent advancements in GAI technology over the seven months since the tests were conducted, and the continuous evolution of AI capabilities could affect the conclusions of this study. Future research should address these biases, incorporate the latest advancements, and consider the dynamic nature of GAI to provide more comprehensive and generalized results.

References

Abba, T. (2024). Use of Artificial Intelligence Technologies in Rendering Library Services: An Empirical Evidence from University Libraries in Africa. African Journal of Library, Archives & Information Science, 34(1), 23–35. https://doi.org/10.4314/ajlais.v34i1.2

Abeywickrama, P., and Mendis, D. (2013). Assessing Sinhala. In The Companion to Language Assessment (pp. 1841–1850). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118411360.wbcla040

Almasre, M. (2024). Development and Evaluation of a Custom GPT for the Assessment of Students' Designs in a Typography Course. Education Sciences, 14(2), Article 2. https://doi.org/10.3390/educsci14020148

Amram, T., Malamud, R. G., & Hollingsworth, C. (2023). Response to" From ChatGPT to CatGPT". Information Technology and Libraries, 42(4). https://ital.corejournals.org/index.php/ital/article/download/16983/11877

Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a Twenty-First Century Library Catalog. Information Technology & Libraries, 25(3), 128–139. https://doi.org/10.6017/ital.v25i3.3342

Bailey, A., & Shou, J. (2021). How Cataloging Can Support Changes and Challenges Affecting Our Libraries. A Report of the ALCTS CaMMS Cataloging and Classification Research Interest Group Virtual Session, ALCTS Virtual Interest Group Week, June 10, 2020. Technical Services Quarterly, 38(1), 88–92. https://doi.org/10.1080/07317131.2020.1854582

Baldwin, P. E. (1976). Bibliographic control problems and organizational change issues posed by the implementation of a computer based cooperative cataloguing support system. https://summit.sfu.ca/_flysystem/fedora/sfu_migrate/2857/b10541159.pdf

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. Future Internet, 15(8), 260. https://doi.org/10.3390/fi15080260

Barbaglia, L., Consoli, S., & Manzan, S. (2021). Exploring the Predictive Power of News and Neural Machine Learning Models for Economic Forecasting. In V. Bitetta, I. Bordino, A. Ferretti, F. Gullo, G. Ponti, & L. Severini (Eds.), Mining Data for Financial Applications (Vol. 12591, pp.

135–149). Springer International Publishing. https://doi.org/10.1007/978-3-030-66981-2_11

Bhatt, M., & Shah, P. (2023). Acceptance of Artificial Intelligence in Human Resource Practices by Employees. In P. Tyagi, N. Chilamkurti, S. Grima, K. Sood, & B. Balusamy (Eds.), The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part B (pp. 13–30). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80455-662-720230002

Breeding, M. (2022). Disruption in the library bibliographic services arena. Library Technology Newsletter, 1(06), 1.

Brzustowicz, R. (2023). From ChatGPT to CatGPT: The implications of artificial intelligence on library cataloging. Information Technology and Libraries, 42(3). https://ital.corejournals.org/index.php/ital/article/view/16295

Cabonero, D. A., Jandoc, J. A. P., Cabalo, J. T., & Dominguez, W. B. (2021). Digging the Issues and Concerns in Subject Access Towards Discoverability of Information and Resources. Library Philosophy and Practice, 1–20.

Chaplin, A. H. (1956). A Universal Cataloging Code. The Library Quarterly, 26(4), 337–347. https://doi.org/10.1086/618347

Fell, S., & Ou, C. (2022). Revisiting Oral History Cataloging Policies for Multi-Format Discovery in MARC: A Case Study. Technical Services Quarterly, 39(3), 225–240. https://doi.org/10.1080/07317131.2022.2082655

Floyd, D. (2023). A response to From ChatGPT to ChatGPT. Information Technology in Libraries, 42(4), 1–3.

Gamage, R., & Wanigasooriya, P. (2023). A Study on the Difference Between AI-Generated and Human-Expert-Generated Default Cataloging Frameworks. Transformation of Libraries in the Digital Era, 51–55.

Gamage, R., Wanigasooriya, P., Hettiarachchi, N., Balasooriya, H., Gagabadaarachchi, V., Wijesinghe, A., & De Silva, C. H. (Eds.). (2024). MARC 21 Descriptive Bibliographic Framework (DBIB): MARC 21 fields for entering data on computer catalogues recommended by the National Library of Sri Lanka. National Library of Sri Lanka. http://www.natlib.lk/pdf/dbib.pdf

Gamage, R., Wanigasooriya, P., Hettiarachchi, N., Balasooriya, H., Gangabadaarachchi, V., Wijesinghe, A., De Silva, C. H., Subasinghe, K., Shiham, M. S. M., & Seneviratne, C. (Eds.). (2020). MARC21 Descriptive Cataloging Framework recommended by National Library of Sri Lanka. National Library and Documentation Centre of Sri Lanka.

Ghassemiazghandi, M. (2024). An Evaluation of ChatGPT's Translation Accuracy Using BLEU Score. Theory and Practice in Language Studies, 14(4), Article 4. https://doi.org/10.17507/tpls.1404.07

Halagatti, M., Gadag, S., Mahantshetti, S., Hiremath, C. V., Tharkude, D., & Banakar, V. (2023). Artificial Intelligence: The New Tool of Disruption in Educational Performance Assessment. In P. Tyagi, S. Grima, K. Sood, B. Balamurugan, E. Özen, & T. Eleftherios (Eds.), Smart Analytics, Artificial Intelligence and Sustainable Performance Management in a Global Digitalised Economy (Vol. 110A, pp. 261–287). Emerald Publishing Limited. https://doi.org/10.1108/S1569-37592023000110A014

Hines, S. S. (2021). Technical Services in the 21st Century. Emerald Publishing Limited. https://www.emerald.com/insight/content/doi/10.1108/S0732-067120210000042016/full/html

Hook, S. (2023, December 28). Can ChatGPT query a database? PC Guide. https://www.pcguide.com/ai/chatgpt-query-database/

Joint Steering Committee for Development of RDA, American Library Association, Library Association of Australia, Cataloguing Committee, British Library, Canadian Committee on Cataloguing, Library of Congress, Canadian Library Association, Chartered Institute of Library and Information Professionals (Great Britain), & Deutsche Nationalbibliothek. (2015). RDA: Resource Description & Access.

Joint Steering Committee for Revision of AACR & American Library Association. (2002). Anglo-American cataloguing rules. Canadian Library Association ; American Library Association.

Ketheeswaren, S. (2017). Editing author main entry towards complying standards: Study on original catalogue of Jaffna public library (2). 20(2), Article 2. https://doi.org/10.4038/jula.v20i2.7900

Kushwah, S. S., & Singh, R. (2022). Issues and Challenges in Indian Multi-Lingual and Multi Scripts Bibliographic Retrieval Systems. Library Philosophy & Practice, 1–16.

Langchain tutorial. Query a database with OpenAI's ChatGPT. (2023, September 18). [Video recording]. https://www.youtube.com/watch?v=e8HMHLr4Nk0

Lowagie, H. (2023). One automatic cataloging flow: Tests and first results. http://repository.ifla.org/handle/123456789/2686

Malhotra, G., & Ramalingam, M. (2023). Perceived anthropomorphism and purchase intention using artificial intelligence technology: Examining the moderated effect of trust. Journal of Enterprise Information Management, ahead-of-print(ahead-of-print). https://doi.org/10.1108/JEIM-09-2022-0316

Management Solutions. (2018). Machine Learning, a key component in business model transformation. Management Solutions. https://www.managementsolutions.com/sites/default/files/publicaciones/eng/machine-learning.pdf

MARC 21 Format for Bibliographic Data: Table of Contents (Network Development and MARC Standards Office, Library of Congress). (2023, December 7). Library of Congress. https://www.loc.gov/marc/bibliographic/

Mer, A. (2023). Artificial Intelligence in Human Resource Management: Recent Trends and Research Agenda. In S. Grima, E. Thalassinos, G. Georgiana Noja, T. V. Stamataopoulos, T. Vasiljeva, & T. Volkova (Eds.), Digital Transformation, Strategic Resilience, Cyber Security and Risk Management (Vol. 111B, pp. 31–56). Emerald Publishing Limited. https://doi.org/10.1108/S1569-37592023000111B003

Metz, C. (2024, January 8). Robots Learn, Chatbots Visualize: How 2024 Will Be A.I.'s 'Leap Forward.' The New York Times. https://www.nytimes.com/2024/01/08/technology/ai-robots-chatbots-2024.html

Neatrour, A., & Hebron, T. K. (2023). Understanding Linked Data and the Potential for Enhanced Discoverability. In Discoverability in Digital Repositories. Routledge.

Olson, L. (2014, October 1). Cataloging a world of languages. https://litablog.org/2014/10/cataloging-a-world-of-languages/

OpenAI. (2024a). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK1: Extract Bibliographic Details from the Web by ISBN. https://chat.openai.com/share/34cd60c7-2e6b-4bcc-9473-95716c356286

OpenAI. (2024b). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK2: Extract Bibliographic Details from the WorldCat Database by Title and ISBN. https://chat.openai.com/share/47d3733d-1a50-4b87-9865-a0f22071496c

OpenAI. (2024c). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK3: Create MARC21 Catalog Entries from the Attached Cover Images [No Sharable URL]. https://chat.openai.com

OpenAI. (2024d). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK4: Convert Book Cover and/or Title Page Details to MARC Record in Other Foreign Language Materials [No Sharable URL]. https://chat.openai.com

OpenAI. (2024e). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK5: Convert Catalog Card Images to MARC Records [No Sharable URL]. https://chat.openai.com

OpenAI. (2024f). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK6: Convert National Bibliography Page to MARC Record [No Sharable URL]. https://chat.openai.com

OpenAI. (2024g). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK7: Create DBIB Compliant MARC21 Records (The Summary Table of DBIB Framework for Monographs, Serials, and Theses & Dissertations Was given to ChatGPT. Thereafter It Was Asked to Convert Records Complying with the Framework). https://chat.openai.com

OpenAI. (2024h). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK8: Create RDA Compliant MARC21 Records with CATMELK (Only Image Prompts Required). https://chat.openai.com

OpenAI. (2024i). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK9: Display MARC21 Records from File. https://chat.openai.com/share/2013e297-10db-4a4b-a98a-088b2a821e42

OpenAI. (2024j). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK10: Existing MARC Record Error Correction 1. https://chat.openai.com/share/3ea22a87-109e-40c4-be97-61e54195e23c

OpenAI. (2024k). ChatGPT 4 (25.01.2024) [Large Language Model]. TASK10: Existing MARC Record Error Correction 2. https://chat.openai.com/share/5fbf3779-46e6-4784-8b72-d4f176cad2e8

Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. Journal

of Business & Industrial Marketing, 34(7), 1410–1419. https://doi.org/10.1108/JBIM-10-2018-0295

Perry, J. K. (2023, September 21). Bibliographic Records: Academic Librarian Cataloging Explained. Sarah Long. https://www.sarahlong.org/bibliographic-records/

Sampath, M. I. G. S. (2011). The need for library collaboration in public library automation in Sri Lanka = ශ්‍රී ලංකාවේ මහජන පුස්තකාල ස්වයංක්‍රීයකරණයේ දී පුස්තකාල සහයෝගීතාවයේ අවශ්‍යතාවය. Proceedings of the 12th Annual Research Symposium, University of Kelaniya. 12th Annual Research Symposium, University of Kelaniya, Kelaniya. http://repository.kln.ac.lk/handle/123456789/8098

Seikel, M., & Steele, T. (2020). Comparison of Key Entities Within Bibliographic Conceptual Models and Implementations: Definitions, Evolution, and Relationships. Library Resources & Technical Services, 64(2), Article 2. https://doi.org/10.5860/lrts.64n2.62

Singh, A., & Chouhan, T. (2023). Artificial Intelligence in HRM: Role of Emotional–Social Intelligence and Future Work Skill. In P. Tyagi, N. Chilamkurti, S. Grima, K. Sood, & B. Balusamy (Eds.), The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A (pp. 175–196). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80382-027-920231009

Tamil. (n.d.). Encyclopedia Britannica. Retrieved May 25, 2024, from https://www.britannica.com/topic/Tamil

Thompson, J., & Hoover, S. (2023). Discoverability Within the Library: Integrated Systems and Discovery Layers. In Discoverability in Digital Repositories. Routledge.

Top 5 AI chatbots that you need to try: ChatGPT, Google Bard, and more. (2023, November 22). HT Tech. https://tech.hindustantimes.com/tech/news/top-5-ai-chatbots-that-you-need-to-try-chatgpt-google-bard-and-more-71700651528839.html

Tosaka, Y., & Park, J. (2018). Continuing Education in New Standards and Technologies for the Organization of Data and Information: A Report on the Cataloging and Metadata Professional Development Survey. Library Resources & Technical Services, 62(1), 4–15. https://doi.org/10.5860/lrts.62n1.4

Varghese, R., Deshpande, A., Digholkar, G., & Kumar, D. (2023). Deciphering the Role of Artificial Intelligence in Health Care, Learning and Development. In P. Tyagi, N. Chilamkurti, S. Grima, K. Sood, & B. Balusamy (Eds.), The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part B (pp. 149–179). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80455-662-720230010

Vasantharajan, C., Tharmalingam, L., & Thayasivam, U. (2021). Adapting the Tesseract Open-Source OCR Engine for Tamil and Sinhala Legacy Fonts and Creating a Parallel Corpus for Tamil-Sinhala-English. https://doi.org/10.1109/IALP57159.2022.9961304

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Quantitative Science Studies, 2(1), 20–41. https://doi.org/10.1162/qss_a_00112

Wanigasooriya, P., & Gamage, R. (2024, September 6). A syntactic analysis of the accuracy and consistency of catalog records in selected Sri Lankan university OPACs. Information for Development: A Way Forward to Address Current Global Challenges. 3rd International Research Conference of National Library of Sri Lanka (ICNATLIB 2024), Colombo.

Woolcott, L., Payant, A., Skeen, B., & Daybell, P. (2021). Missing the MARC: Utilization of MARC Fields in the Search Process. Cataloging & Classification Quarterly, 59(1), 28–52. https://doi.org/10.1080/01639374.2021.1881010

Xu, R., Feng, Y., & Chen, H. (2023). ChatGPT vs. Google: A comparative study of search performance and user experience. arXiv Preprint arXiv:2307.01135. https://arxiv.org/abs/2307.01135

Zhao, B. Z. H., Ikram, M., & Kaafar, M. A. (2024). GPTs Window Shopping: An analysis of the Landscape of Custom ChatGPT Models (arXiv:2405.10547). arXiv. http://arxiv.org/abs/2405.10547

Zihni, E., McGarry, B. L., & Kelleher, J. D. (2022). Moving Toward Explainable Decisions of Artificial Intelligence Models for the Prediction of Functional Outcomes of Ischemic Stroke Patients. In S. L. Linwood (Ed.), Digital Health. Exon Publications. http://www.ncbi.nlm.nih.gov/books/NBK580624/