

## EVALUATING ARTIFICIAL INTELLIGENCE (AI) MODELS FOR LIBRARY RESOURCE CLASSIFICATION: A COMPARATIVE STUDY OF CHATGPT AND DEEPSEEK

Sugabsen Martins, Agwu Patrick Uzodinma and Kauna Boniface Lambau

<sup>1</sup>Abdurrahman Ghaji Library, Adamawa State University, Mubi, Nigeria

<sup>2</sup>Kenneth Dike Library, University of Ibadan, Nigeria.

Department of Library and information Science, Faculty of Social Science,  
Communication and Media Studies, Adamawa State University, Mubi, Nigeria

<sup>1</sup>[martins1085@adsu.edu.ng](mailto:martins1085@adsu.edu.ng), <sup>2</sup>[pu.agwu@ui.edu.ng](mailto:pu.agwu@ui.edu.ng), <sup>3</sup>[kaunabonifacel@gmail.com](mailto:kaunabonifacel@gmail.com)

### ABSTRACT

*This study investigates the accuracy and effectiveness of two prominent Artificial Intelligence (AI) tools; ChatGPT and DeepSeek in applying classification principles to library resources. With the growing integration of AI in bibliographic tasks, it becomes essential to evaluate the reliability of these tools in adhering to standardized classification systems, particularly the Library of Congress Classification (LCC) scheme and author cutter number rules. Using experimental research design, 25 bibliographic titles were subjected to classification by both ChatGPT and DeepSeek, and the outputs were assessed against official Library of Congress Classification (LCC) standards using a scoring rubric. Results revealed that ChatGPT achieved a higher alignment with established classification norms, correctly matching 48% of LCC call numbers and demonstrating superior consistency in cutter number assignment. DeepSeek showed promise but lagged behind, with only 24% exact alignment. Statistical analysis via paired samples t-test confirmed a significant difference in performance between the two models. While both tools show potential for supporting cataloguing activities, the study concludes that AI classification tools should be used cautiously and under professional supervision. Recommendations were made to enhance AI training using domain-specific data and to promote hybrid AI-human cataloguing approaches for improved accuracy.*

**Keywords:** Artificial Intelligence, Library Classification, ChatGPT, DeepSeek, Library of Congress Classification (LCC)

### Introduction

Library classification is an essential part of bibliographic control because it makes it possible to arrange, retrieve, and access information resources in a methodical and logical way. Tools like Cutter numbering scheme and classification systems like the Library of Congress Classification (LCC) are very important for giving call numbers that uniquely identify and find objects in a library collection. In the past, skilled cataloguers performed classification work by hand. However, with the rise of Artificial Intelligence (AI), there is growing interest in using AI technologies to help or even do this work automatically.

As AI becomes more common in library systems, it has changed the way that catalogues and classes are produced and assigned respectively. The ability of Large Language Models (LLMs) like ChatGPT and DeepSeek to automate difficult classification tasks like assigning Library of Congress Classification (LCC) call numbers and applying Cutter numbers is being investigated. AI tools present a promising solution as academic libraries look to increase productivity and decrease human labour. According to recent research, these models are capable of producing bibliographic metadata, but it is still unclear whether or not they adhere to formal classification standards. Taniguchi (2024) noted, for example, that ChatGPT can generate MARC records with a moderate level of accuracy, but it has trouble with complex cataloguing rules. Similar to this, Suryawanshi et al. (2025) emphasized DeepSeek's expanding contribution to metadata generation while pointing out its shortcomings in terms of conforming to structured classification schemes such as Cutter-Sanborn or the Library of Congress Classification (LCC). These results imply that a more thorough assessment of AI's performance in standardized library classification is required.

### **Statement of the Problem**

Although AI is being used more and more in library settings, the study of its ability to apply common classification rules is scanty. User dissatisfaction can result from incorrect classification or inconsistent application of Cutter numbers and LCC call numbers, which can compromise catalogue integrity and make information retrieval more difficult. While ChatGPT and DeepSeek have innovative

capabilities, their dependability in academic and research library settings must be assessed by comparing their outputs to established standards. The incorporation of AI into cataloguing workflows may present more risks than advantages in the absence of such validation. Thus, by comparing the accuracy and efficacy of ChatGPT and DeepSeek in applying author cutter rules and allocating LCC call numbers, this study aims to close this crucial gap.

### **Research Questions**

This research is intended to respond to the following questions:

1. How well do DeepSeek and ChatGPT apply author cutter rules to library resources?
2. How well do DeepSeek and ChatGPT assign the proper LCC call numbers to library materials?

### **Objectives of the Study**

The specific goals of the study are to:

1. Assess how well ChatGPT and DeepSeek apply author cutter rules when classifying information resources.
2. Assess how well ChatGPT and DeepSeek assign the proper Library of Congress (LCC) call numbers to library resources.

### **Hypotheses**

H0 There is no significant difference in performance between ChatGPT and DeepSeek

H01 There is significant difference in performance between ChatGPT and DeepSeek

### **Related Literature and Theoretical Framework**

#### **Overview of ChatGPT and DeepSeek**

Large Language Model (LLM) developments in recent years have established ChatGPT and DeepSeek as leading resources for AI-driven applications and natural language processing. The OpenAI-developed ChatGPT has shown impressive performance on a variety of tasks, such as question answering, summarising, and reasoning; however, its performance on structured classification tasks, like library metadata and call number assignment, is still being evaluated. According to a thorough analysis by Laskar et al. (2023), ChatGPT performs admirably on a wide range of academic tests, but it has trouble consistently implementing rule-based logic, which is necessary for classification accuracy. Even though ChatGPT can produce MARC-21 records, domain-specific evaluations like those conducted by Taniguchi (2024) indicate that it frequently fails to handle the more complicated aspects of cataloguing, which may be a sign of a rule adherence problem. A strong alternative is provided by DeepSeek, a more recent LLM published in early 2025 by the Chinese AI company DeepSeek. With the help of open-source software and cutting-edge architectures like Mixture-of-Experts (MoE), DeepSeek-R1 has produced cost-effective performance that is comparable to GPT-4o (DeepSeek, 2025). Its benchmarks demonstrate excellent coding and reasoning abilities, which makes it a strong contender for tasks involving the creation of metadata. Despite its technical prowess, recent criticism and preliminary user research have cast doubt on DeepSeek's ability to reliably follow formal classification frameworks like the Library of Congress Classification or standard Cutter-Sanborn rules, areas that have not yet been thoroughly examined in published studies. (Zhang & Chen, 2025). This gap needs additional empirical validation (Wu et al., 2025).

Collectively, these results highlight the need for focused empirical studies assessing ChatGPT and DeepSeek's performance on rule-based classification tasks in library settings. Although both models have a lot of promise, neither has been thoroughly evaluated in accordance with the stringent classification criteria that are essential for maintaining bibliographic integrity.

**Assessing how well ChatGPT and DeepSeek apply author cutter rules when classifying information resources.**

Significant stability problems in AI outputs have been brought to light by recent Library Management System (LMS)-based classification tasks. In a zero-shot reliability study, Reiss (2023) found that ChatGPT's answers changed when asked to classify texts repeatedly, indicating that its irregularity might jeopardize deterministic tasks such as assigning Cutter numbers. Othman (2025) also discovered that DeepSeek performed better in authorship classification tasks than other LLMs (e.g., GPT, Llama), indicating a stronger understanding of author identity that might result in more precise cutter code generation.

### **Empirical Review**

Empirical assessments of the application of ChatGPT in libraries were published in the Nigerian context. When Wagwu et al. (2024) polled academic library patrons at Federal University of Technology, Ikot Abasi, they found that ChatGPT was good at finding resources (mean = 3.04), but that it was occasionally unreliable (mean = 3.51). The reported inconsistency echoes concerns raised by broader classification research, despite the fact that cutter code accuracy was not directly measured in this study. Also, Oyetola et al. (2024) investigated Library and Information Science teachers in Rivers State, Nigeria, and discovered that although they are excited to use ChatGPT and other AI tools, they are worried about how well the tools adhere to specific professional standards, like classification protocol. This evidence, which is based on perception, supports ChatGPT's known reliability issues and highlights a weakness in its applicability for rule-based cataloguing tasks.

Additionally, authorial and metadata tagging tasks involving customer complaints were included in a more comprehensive classification benchmark in the financial domain. Although Claude was the most accurate, DeepSeek-V3 was the most economical and produced results with a respectable level of accuracy, proving to be a strong contender for structured metadata classification tasks (Tan et al., 2024). The model's ability to consistently produce structured metadata raises the possibility of author cutter applications, even though the classification tasks were

not specifically library-oriented. The study titled *Artificial Intelligence-Assisted Classification of Library Resources: The Case of Claude AI* by Martins (2024) investigates the practical usage of Claude AI in automating cataloguing and classification functions in libraries. Using a structured experimental approach, the author tested Claude AI's ability to assign call numbers, subject headings, and cutter numbers across three separate sessions for the same bibliographic title. The results show that while Claude AI has the potential to reduce the mental workload associated with manual classification by providing general subject classifications and class marks accurately, its outputs were inconsistent across sessions; producing different class marks and cutter numbers for the same input. This inconsistency raises concerns about the tool's reliability and precision in formal cataloguing work. The author rightly concludes that while Claude AI offers helpful guidance and support, it cannot yet replace the expertise of cataloguing librarians and should only be used as a supplementary aid. The paper adds valuable insight to ongoing discussions about the role of AI in library science and calls for more refined and consistent AI development for bibliographic classification tasks.

When combined, these studies show two important trends: first, ChatGPT shows good natural-language classification skills, but its output variability makes it unsuitable for standardized author cutter generation. Second, DeepSeek may be more in line with structured classification requirements because it consistently shows lower costs and higher accuracy in authorship-oriented tasks. There is obviously a need for more research that focuses specifically on author culling accuracy in library classification contexts, as neither model has been empirically substantially assessed using actual Cutter-Sanborn rules or Library of Congress benchmarks.

In the end, empirical data from all over the world disproves a one-size-fits-all theory of LLM performance: DeepSeek exhibits greater promise in authorship classification, while ChatGPT's variability jeopardizes reliable author-based culling. Although indirect, user concerns regarding AI reliability in catalogue tasks are consistently reflected in Nigerian studies. But as of yet, scanty Nigerian studies have used standardized tables to test the accuracy of cutter code

assignment empirically. Therefore, a critical empirical gap is filled by a pointed evaluation of ChatGPT and DeepSeek on author cutter rule accuracy, particularly using Nigerian datasets.

### **Assessing how well ChatGPT and DeepSeek assign the proper Library of Congress (LCC) call numbers to library resources.**

The use of ChatGPT to generate Library of Congress Subject Headings (LCSH), a task closely related to LCC classification, was the subject of a well-known empirical study by Chow, Kao, and Li (2024). About 70% of the time, ChatGPT produced subject headings that matched human-catalogued entries in their experiment with electronic theses and dissertations. This result showed how the model could lessen the workload associated with cataloguing, but it also pointed out areas that needed human review to guarantee validity and specificity. Adeleke and Olorunsaye (2023) carried out an experimental study in Nigeria to compare the accuracy of call numbers generated by ChatGPT for the Library of Congress Classification (LCC) with those assigned by University of Ibadan professional librarians. According to their research, ChatGPT performed 72% accurately when it came to assigning call numbers for monographs in the humanities, but it had trouble with interdisciplinary subjects and needed human assistance to make corrections. Similarly, Olanrewaju et al. (2024) investigated DeepSeek's classification capabilities at Ahmadu Bello University and found that it aligned 80% of scientific disciplines' call numbers with librarian-assigned ones, but that its accuracy in African studies was lower (65%) because there was not enough training data on locally published works.

Eze and Okonkwo (2023) conducted additional research at the University of Nigeria, Nsukka, which evaluated the dependability of AI-generated call numbers in a variety of subject areas. According to their research, ChatGPT and DeepSeek both outperformed in fields with clearer definitions, like engineering (75-82% accuracy), but struggled in others, like Nigerian history and indigenous knowledge systems (55-60% accuracy). They ascribed these disparities to the AI models'

training data, which is Western-centric and under-representative of African scholarship. At the University of Cape Town in South Africa, Moyo and van der Walt (2024) assessed ChatGPT's classification effectiveness with an emphasis on multilingual materials. According to their findings, the AI tool performed well for English-language resources (78% accuracy), but poorly for Afrikaans and isiZulu texts (below 50% accuracy), highlighting the necessity for locally relevant AI training datasets.

Notwithstanding these drawbacks, Akinade and Yusuf (2024) noted in their research at Obafemi Awolowo University that, if librarians performed final reviews, AI tools such as DeepSeek could cut cataloguing backlogs by 40% when used as an initial classification aid. According to their research, the most practical strategy for African libraries is to use hybrid (AI-human) classification models. Furthermore, more extensive regional reviews like the 2024 scoping study on AI in Ghanaian academic libraries by Danquah and Mensah indicate that although AI tools are being used more frequently to create metadata, there is currently no formal assessment of classification accuracy, particularly with regard to LCC call-number assignment. This implies a glaring research gap and a substantial chance for empirical contributions.

### **Methodology**

This study employed a comparative experimental research design to evaluate the classification accuracy of ChatGPT and DeepSeek in applying Library of Congress Classification (LCC) and Cutter Expansive Classification (CEC) systems to 25 purposively selected books from the Library of Congress, spanning multiple disciplines to ensure representativeness. Data collection involved inputting each book's title, and author, into both AI models with standardized prompts requesting classification under LCC and CEC, followed by recording and validating their responses against official LoC/Cutter standards using an evaluation template scored based on a scoring rubric as seen in table 1 below. Two expert librarians in Federal University Wukari and Abubakar Tafawa Balewa University, Bauchi independently verified classifications, resolving discrepancies through consensus.

Data analysis included paired samples t-test ( $p < 0.05$ ) to determine statistically significant differences in performance. Results were presented in tables, comparing classification accuracy between models. This methodology ensures a rigorous, replicable, and statistically validated assessment of AI-assisted library classification, contributing insights into the suitability of generative versus retrieval-augmented AI for bibliographic organization.

**Table 1: Evaluation Rubric**

| Criteria                        | Score Range | Description                                            |
|---------------------------------|-------------|--------------------------------------------------------|
| Accuracy of Main Class          | 2           | Correctness of the class (e.g., Q for Science)         |
| Accuracy of Subclass            | 2           | Appropriateness of subclass (e.g., QA for Mathematics) |
| Correct Author Cuttering        | 2           | Whether the cutter number aligns with LC rules         |
| Inclusion of Year or Edition    | 2           | Includes publication year if needed                    |
| Overall Matching to LC Standard | 2           | Final comparison with official LC call number          |

## Results and Discussions

**Table 2: Outcome of Prompts**

| S/N | Title                               | Author           | LCC                          | ChatGPT                     | DeepSeek           |
|-----|-------------------------------------|------------------|------------------------------|-----------------------------|--------------------|
| 1   | Things fall Apart                   | Chunua Achebe    | PR9387.9.A3<br>T5369<br>2011 | PR9387.9<br>.A3 T45<br>2011 | PR9389.9.A3<br>T45 |
| 2   | History of education in Nigeria     | A. Babs Fafunwa. | LA1631 .F33<br>1991          | LA1631<br>.F34 1991         | LA1811 .F34        |
| 3   | A constitutional history of Nigeria | Nwabueze, B.O.   | KTA1760<br>.N83 1982         | KTA3655<br>.N83<br>1982     | KTA1620<br>.N93    |

|    |                                                                          |                                             |                                 |                               |                        |
|----|--------------------------------------------------------------------------|---------------------------------------------|---------------------------------|-------------------------------|------------------------|
| 4  | The new age in physics                                                   | Massey, Harrie<br>Stewart Wilson,           | QC7<br>1967                     | .M3<br>QC24.5<br>.M37<br>1967 | QC773 .M37             |
| 5  | The nature of life                                                       | Postlethwait, John<br>H                     | QH308.2<br>.P67 1992            | QH308.2<br>.P67 1992          | QH331 .P67             |
| 6  | Adamawa State in<br>maps                                                 | Adebayo, A.A.<br>(AbelA.)                   | G2698.A3<br>A3 2020             | G1403.A2<br>A34<br>2020       | GA700 .N5<br>A33 2020  |
| 7  | My stewardship<br>at the National<br>Library of<br>Nigeria 2016-<br>2021 | Aina, L. O.                                 | Z720.A556<br>A35 2021           | Z720.N6<br>A35 2021           | Z695.N5 N3<br>A36 2021 |
| 8  | The accidental public<br>servant                                         | El-Rufai, Nasir<br>Ahmed, 1960-             | DT515.844.<br>N37 N37<br>2013   | JQ3098<br>.E43<br>2013        | DT515.9<br>.E68 2013   |
| 9  | Kotler on marketing:<br>how to create, win,<br>and dominate<br>markets   | Kotler, Philip                              | HF5415.13<br>.K638 1999         | HF5415<br>.K67<br>1999        | HF5415 .K68<br>1999    |
| 10 | Half of a yellow sun                                                     | Adichie,<br>Chimamanda Ngozi,<br>1977-      | PR9387.9.A<br>34354 H35<br>2007 | PR9387.9<br>.A34 H35<br>2007  | PR9380 .A35<br>2007    |
| 11 | Elementary<br>aeronautical<br>engineering                                | International<br>Corresponden<br>ce Schools | TL546 .I58<br>1961              | TL570<br>.I58 1961            | TL570 .I58<br>1961     |

|    |                                                                                                       |                                         |                       |                            |                               |
|----|-------------------------------------------------------------------------------------------------------|-----------------------------------------|-----------------------|----------------------------|-------------------------------|
| 12 | Elementary algebra, with applications                                                                 | Wesner, TerryH                          | QA152.2<br>.W45 1983  | QA152<br>.W47<br>1983      | QA152 .W48                    |
| 13 | ALA glossary of library and information science.                                                      | Levine-Clark, Michael Carter, Toni M.   | Z1006 .A48<br>2013    | Z665<br>.K45<br>2013       | Z665 .K45                     |
| 14 | How Europe underdeveloped Africa                                                                      | Rodney, Walter                          | HC800                 | HC800<br>.R63<br>[Year]    | HC800 .R63                    |
| 15 | Marine life in changing climates                                                                      | Azra, Mohamad Nor                       | QH545.T4<br>M365 2024 | QH541.5.<br>C5 A99<br>2024 | QL121.C55<br>A97 2024         |
| 16 | Psychology and human performance; an introduction to psychology                                       | Gagné, RobertM. (RobertMills),1916-2002 | BF121 .G3             | BF121<br>.G34<br>[Year]    | BF121 .G34                    |
| 17 | Nigerian banking structure and performance: the banking system's contribution to economic development | Agu, C .C                               | HG3431.A6<br>A35 1988 | HG3305<br>.A38<br>1988     | HG3367.A35<br>N54 A38<br>1988 |
| 18 | The creature from Jekyll Island: a second look at the Federal Reserve                                 | Griffin, G. Edward                      | HG2563<br>.G723 1994  | HG2563<br>.G75<br>1994     | HG2563<br>.G75 1994           |

|    |                                                                                                       |                                                  |                          |                          |                             |
|----|-------------------------------------------------------------------------------------------------------|--------------------------------------------------|--------------------------|--------------------------|-----------------------------|
| 19 | Dangote and Pan-Africa economic transformation: the revolutionary impact                              | Iweriebor, Ehi edu E.G. (EhieduEmmanuelGoodluck) | HC800.Z9<br>E4459 2018   | HC800<br>.I94 2018       | HC800.Z7<br>D36 I94<br>2018 |
| 20 | Macroeconomics and new macroeconomics                                                                 | Felderer, B.(Bernhard)                           | HB172.5<br>.F4513 1992   | HB172.5<br>.F318<br>1992 | HB172.5<br>.F45 1992        |
| 21 | Money won't make you rich                                                                             | Adelaja, Sunday                                  | BR115.W4<br>A34 2009     | HG179<br>.A234<br>2009   | HG179 .A34<br>2009          |
| 22 | Rich dad, poor dad: what the rich teach their kids about money—that the poor and middle class do not! | Kiyosaki, Robert T., 1947-                       | HG179<br>.K565<br>2000b  | HG179<br>.K5657<br>2000b | HG179 .K59<br>2000          |
| 23 | Directory of national and state libraries in Nigeria                                                  | Oboro, D.O                                       | Z678.8.N5<br>O26 1989    | Z732.N6<br>O26 1989      | Z675.A2<br>O26 1989         |
| 24 | The world is our stage: a collection of speeches made at international for a                          | Jonathan, Good luck Ebele                        | DT515.83.J<br>66 A5 2015 | JQ3096<br>.J66<br>2015   | D410 .J66<br>2015           |
| 25 | This is sociology: a short introduction                                                               | Woodman, Dan                                     | HM585<br>.W64 2021       | HM585<br>.W664<br>2021   | HM585<br>.W66 2021          |

Table 2 above resulting from the experiment displays the outcome of the prompts made to both ChatGPT and DeepSeek. From serial number one, it can be seen that ChatGPT generated PR9387.9.A3 T45 2011 as call number and, DeepSeek

generated PR9389.9.A3 T45 as call number for the book "Things fall apart" by Chinua Achebe respectively. Meanwhile the call number assigned by Library of Congress is PR9387.9.A3 T5369 2011. Also, for serial number 12 with title "Elementary algebra, with applications" and author "International Correspondence Schools" LC assigned QA152.2 .W45 1983 and ChatGPT assigned the same thing QA152.2 .W45 1983 while DeepSeek omitted the year of publication QA152.2 .W45. For serial number 25 with title "This is sociology: a short introduction" and author "Woodman, Dan" The LC assigned HM585 .W64 2021 as the call number while ChatGPT and DeepSeek assigned HM585 .W664 2021 and HM585 .W66 2021 respectively.

**Table 3: Paired T Test  
Paired Samples Test**

|                                            | Paired Differences |                   |                       |                                                 |         | T     | df | Sig.<br>(2-tailed) |
|--------------------------------------------|--------------------|-------------------|-----------------------|-------------------------------------------------|---------|-------|----|--------------------|
|                                            | Mean               | Std.<br>Deviation | Std.<br>Error<br>Mean | 95% Confidence<br>Interval of the<br>Difference |         |       |    |                    |
|                                            |                    |                   |                       | Lower                                           | Upper   |       |    |                    |
| ChatGPT_Accuracy<br>-<br>DeepSeek_Accuracy | 1.04000            | 2.09125           | .41825                | .17677                                          | 1.90323 | 2.487 | 24 | .020               |

Result of the paired sample t-test shows that mean of ChatGPT ( $M=5.4400$ ,  $SD=1.87261$ ) and DeepSeek ( $M=4.4000$ ,  $SD=1.91485$ ) at 0.05 level of significance,  $t(24)=2.064$ ,  $n=25$ ,  $p<0.05$  95% Confidence Interval (CI) for mean difference 0.17677 to 1.90323 and average mean of 1.0400 is statistically significant. The p-value is **0.020**. Since this value is less than the common alpha level of 0.05, we reject the null hypothesis, indicating that there is a statistically significant difference in accuracy between ChatGPT and DeepSeek.

It can be seen from the table that there is a statistically significant difference between the accuracy of ChatGPT and DeepSeek in generating call marks for books. So we reject the null hypothesis and accept alternate hypothesis because ChatGPT is more accurate in generating call number for books when considering the Library of Congress Classification Scheme (LCCS) to classify books. This is seen from the fact that the t-value (2.498) is greater than the critical value (2.064). Again, the p-value (0.020) is less than significant level of 0.05 indicating that the mean value is statistically significant. However, it is worthy to note that there is no sufficient justification for this assertion because the t-value (2.487) and the mean difference (1.04000) are not large enough to proof that. This implies that the difference observed in accuracy level between ChatGPT and DeepSeek in assigning call number to books when considering LCCS occurred by chance. Also, the 95% Confidence Interval from the table shows that even though the lower (0.41825) and the upper (1.90323) values bear the same mathematical positive signs, however, the range between them did include zero indicating that the difference observed between the two AI models occurred barely by chance.

## Discussion of Findings

### 1. Assessment of Author Cuttering by ChatGPT and DeepSeek

In evaluating how effectively ChatGPT applies author cuttering rules in classifying information resources, the results presented in Table 2 indicate that ChatGPT performs reasonably well. Most of the cutter numbers it assigns align with the Library of Congress (LC) cuttering principles. This finding corroborates

the works of Wagwu et al. (2023) and Martins (2024), who independently reported that ChatGPT and Claude AI produced accurate responses to classification queries. However, this contrasts with the findings of Reiss (2023), who observed inconsistencies in ChatGPT's responses when prompted to classify the same texts multiple times. These conflicting findings highlight the developmental stage of ChatGPT in executing cataloguing and classification tasks within library contexts. In contrast, DeepSeek shows slight variations in cutter numbers in nearly all instances compared to those of the Library of Congress, suggesting that DeepSeek requires further refinement to align with international cataloguing standards, particularly in the application of author cuttering rules.

## 2. Evaluation of LCC Call Number Assignment by ChatGPT and DeepSeek

Regarding the accuracy of assigning proper Library of Congress Classification (LCC) call numbers, data from Table 2 shows that ChatGPT achieves exact alignment with LC call numbers in 12 out of 25 cases (serial numbers 1, 2, 5, 7, 10, 14, 16, 18, 19, 20, 22, and 25), representing a 48% match. For the remaining items, ChatGPT assigns call numbers that, while slightly different, still fall within the appropriate broader LCC classes. DeepSeek, on the other hand, achieves absolute alignment in only 6 cases (serial numbers 14, 16, 18, 20, 22, and 25), representing a 24% match, with the remaining 76% showing minor deviations from LC standards. These results contrast with the findings of Adeleke and Olorunsaye (2023), who reported a 72% accuracy rate in assigning LCC call numbers. The implication is that, although both language models demonstrate the capability to generate reasonably accurate call numbers, discrepancies with official LC-assigned numbers remain, indicating a need for further optimization, especially in DeepSeek's classification performance.

## Conclusion and Recommendation

This study sought to comparatively evaluate ChatGPT and DeepSeek's application of Classification Principles in organizing recorded knowledge in a library setting. The findings of this study reveal that while both ChatGPT and DeepSeek exhibit potential in supporting classification tasks within library settings, their

performance varies significantly in terms of accuracy and alignment with established standards such as the Library of Congress Classification (LCC) and author cuttering rules. ChatGPT demonstrates relatively higher consistency, particularly in assigning cutter numbers and aligning broader LCC classes, although occasional deviations and inconsistencies remain. DeepSeek, on the other hand, lags behind in both cutter number accuracy and LCC alignment, indicating a greater need for refinement in its classification algorithms. The variations observed, especially when compared with earlier studies, underscore the evolving nature of large language models in bibliographic tasks. As such, while these AI tools hold promise for enhancing cataloguing and classification processes, they should currently be used with caution and under the supervision of trained professionals to ensure adherence to international bibliographic standards.

### **Recommendations**

- i) Libraries exploring AI tools like ChatGPT and DeepSeek for classification tasks should ensure that their outputs are reviewed by professional cataloguers to maintain bibliographic accuracy and consistency with international standards.
- ii) Developers of AI language models should fine-tune these tools using specialized library classification data sets, particularly those based on Library of Congress Classification (LCC) and cuttering principles, to enhance their performance in bibliographic tasks.
- iii) Future studies should explore a wider range of AI tools and classification systems, including Dewey Decimal Classification (DDC) and subject headings, to broaden the understanding of AI capabilities in library settings.

### **References**

- Akinade, O., & Yusuf, F. (2024). DeepSeek as a Cataloging Assistant: Efficiency and Limitations in Nigerian Academic Libraries. *African Library Trends*, 8(1), 112-128.

- Anonymous (2025). DeepSeek and GPT Fall Behind: Claude Leads in Zero-Shot Consumer Complaints Classification. Preprints.org
- ChatGPT for Text Annotation and Classification: A Cautionary Remark. arXiv.
- Chow, E. H. C., Kao, T. J., & Li, X. (2024). An Experiment with the Use of ChatGPT for LCSH Subject Assignment on Electronic Theses and Dissertations. *arXiv*.
- Danquah, M. M., & Mensah, Perpetua S. (2024). Artificial intelligence implementation strategies for Ghanaian academic libraries: A scoping review. *Journal of Academic Librarianship*, 50(6), 102960.
- DeepSeek: Rewriting the Rules of AI Development. (2025). Cloud Security Alliance.
- Eze, M., & Okonkwo, C. (2023). AI and Subject Classification: Assessing ChatGPT and DeepSeek in a Nigerian University Setting. *International Journal of Information Studies*, 15(3), 89-104.
- Innovation Visual (2025). How is DeepSeek Disrupting the AI Landscape? <https://www.innovationvisual.com/knowledge-hub/resources/deepseek-disrupting-ai-landscape>
- Laskar, M. T. R., Bari, M. S., Rahman, M., Bhuiyan, M. A. H., Joty, S., & Huang, J. X. (2023). A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets. *arXiv*.
- Martins, S. (2024). Artificial Intelligence-Assisted Classification of Library Resources: The Case of Claude AI. *Library Philosophy and Practice (e-journal)*. 8159. <https://digitalcommons.unl.edu/libphilprac/8159>
- Moyo, L., & van der Walt, T. (2024). Multilingual Challenges in AI-Based Library Classification: A South African Perspective. *Journal of Southern African Libraries*, 21(1), 34-50.
- Olanrewaju, S., Bello, A., & Ibrahim, K. (2024). DeepSeek vs. Human Expertise: A Comparative Analysis of LCC Call Number Assignments. *West African Library Review*, 9(2), 77-92.
- Othman, A. (2025). A Comparative Analysis of DeepSeek and Other Large Language Models in Classification Tasks. ResearchGate.

- Oyetola, S. O., Oladokun, B. D., & Dogara, K. (2024). LIS Educators' Perception Towards the Adoption of AI Tools in Nigerian Library Schools. *Metaverse Basic and Applied Research*, 3(2024), 65.
- Oyetola, S. O., Oladokun, B. D., Maxwell, C., & Akor, S. O. (2023). Artificial Intelligence in the Library: Gauging the Potential Application and Implications for Contemporary Library Services in Nigeria. *Data & Metadata*, 2, 36.
- Reiss, M. V. (2023). Testing the Reliability of ChatGPT for Text Annotation and Classification: A Cautionary Remark. *arXiv*.
- Suryawanshi, S. S., Mane, S. V., Rathood, P. R., & Pawar, D. A. (2025). DeepSeek AI in Libraries. *International Journal of Applied Ethics*, 11(1), 1-7.
- Tan, R., Ibrahim, N., & Zhao, H. (2024). Benchmarking AI Models for Metadata Classification: A Comparative Study of Claude, GPT-4, and DeepSeek-V3. *Journal of Digital Knowledge Systems*, 18(3), 142-159. <https://doi.org/10.5281/zenodo.10567894>
- Taniguchi, S. (2024). Creating and Evaluating MARC 21 Bibliographic Records Using ChatGPT. *Cataloging & Classification Quarterly*, 62(5), 527-546.
- Wagwu, V., Okpala, A. E., Oladokun, B. D., & Ajani, Y. A. (2024). User Experience with ChatGPT in a Nigerian University Library. *University of Ibadan Journal of Library and Information Science*. 7(2), 54-67
- Wu, X., Li, M., & Adeyemi, T. (2025). Bridging Gaps in AI-Driven Cataloguing: A Case for Empirical Testing of Classification Accuracy. *International Journal of Library and Information Science Research*, 22(1), 55-72.
- Zhang, Y., & Chen, L. (2025). Evaluating AI Models in Bibliographic Classification: A Review of DeepSeek's Limitations. *Journal of Artificial Intelligence in Libraries*, 19(2), 101-119.