Evaluating AI-Generated Subject Terms for Cataloging Assistance

Li Sun*

This study evaluated AI-generated faceted application of subject terminology (FAST) subject headings using only call numbers, titles, and authors from 212,000 brief MARC records. A stratified sample of 1,000 records across five domains was analyzed. AI terms were compared to OCLC FAST headings through direct and enhanced (semantic) matching. Enhanced matching outperformed direct matching, with F1 scores highest in history (0.81) and lowest in sciences (0.72) due to specialized terminology. AI processed large datasets efficiently but produced broad terms. Confidence scores strongly predicted match quality ($r = 0.78$, $p < 0.01$) for enhanced matching, supporting their use in a hybrid AI–human workflow. Future work should focus on domain-specific training.

Introduction

Artificial intelligence (AI) has demonstrated considerable potential in automating cataloging processes. Several library professionals have explored the use of generative AI tools to create Library of Congress Subject Headings (LCSH) based on titles and abstracts (Chow et al., 2024). However, no research has yet examined the use of more concise metadata for this task. Specifically, there is a gap in the literature regarding the application of AI to generate subject terms based solely on call numbers, titles, and authors. Addressing this gap is both necessary and timely. At the author's library, a large collection of items has brief MARC records with only those three fields. Without subject terms, users cannot easily find these items unless they know the exact title or browse the shelves. To improve access, these records need richer metadata, especially subject headings. However manually updating thousands of records is not feasible, and automation becomes the optimal solution. This study explores whether a generative AI tool can produce consistent, standardized subject terms using its built-in models and vocabularies.

During the COVID-19 lockdown, staff worked remotely. The cataloging librarian created remote projects and used Alma Analytics to find records needing attention. Many were legacy entries from the shift to digital systems in the 1970s and 1980s. They included only basic bibliographic data. In total, 212,000 brief records were found, including 68,000 with Library of Congress Control Numbers (LCCNs). These LCCNs were used to retrieve full MARC records from OCLC and overlay the existing brief records. That process was slow but doable.

*Li Sun is Head of the East Asian Languages Section and Head of the Continuing Resources Section at the Rutgers University Libraries, email: lisun@rutgers.edu. ©2026 Li Sun, Attribution-NonCommercial (https://creativecommons.org/licenses/by-nc/4.0/) CC BY-NC.

Once the LCCNs were exhausted, addressing the remaining brief records became impractical. This study focuses on records without LCCNs and other identifiers, where OCLC retrieval is not feasible, precluding direct comparison with copy-cataloging workflows. The author turned to AI automation to define subject terms. To ensure that the AI performs as expected, thorough testing was conducted. Initial tests and past research show AI can generate broad subject terms but struggles with LCSH formatting. This study aims to evaluate the accuracy and appropriateness of AI-generated subjects by comparing them with faceted application of subject terminology (FAST) subject headings. These FAST subject terms are incorporated into OCLC's MARC records within the 6xx field with a second indicator of 7, distinguishing them from LCSH. Details of the testing design are described in the Methods section.

Literature Review

AI for Subject Heading Generation

Recent studies have explored generative AI's role in automating subject heading assignment. Studies consistently show that AI enhances cataloging efficiency; however, these studies also stress the need for human oversight for accuracy, completeness, and relevance. Among others, Chow et al. (2024) provided ChatGPT with titles and abstracts from electronic theses and dissertations. They found that AI streamlined LCSH generation and improved resource accessibility. Still, verification remained essential. Similarly, Martins (2024) tested Anthropic's Claude AI on a single book to assign call numbers, subject terms, and cutter numbers. While general subjects proved relevant, specific indexing was inconsistent across sessions. This reinforced AI's role as a supplement, not a replacement, despite the study's narrow scope. Zhang et al. (2023) complemented these efforts. They used prompt-based training on large language models. Input came from dataset titles and descriptions in Research Data Australia. Results showed strong automatic subject annotations. But accuracy fell when discipline-specific rules needed more context. Earlier machine-learning approaches laid foundational groundwork. For example, Frank and Paynter (2004) mapped LCSH to classifications in a reverse process; Mödden (2022) automated Dewey Decimal assignments via shorter classifications. Collectively, these works highlight AI's transformative value in deriving subject terms from limited inputs when paired with human expertise in cataloging efforts. These studies suggest that, while certain limitations persist, generative AI has proven effective in deriving subject terms from input data.

Building on these studies, this present research examines the challenges posed by incomplete MARC records and evaluates the ability of AI models to generate standardized subject terms aligned with the FAST framework. FAST is a modern, simplified version of the complex LCSH. It was designed to make subject cataloging and searching easier (OCLC, 2025). Created by OCLC in collaboration with the Library of Congress, FAST separates subject information into nine distinct, independent facets, such as topical, personal names, and chronological. This "post-coordinated" approach replaces the long, single strings of traditional headings, allowing users to combine facets to refine their searches, much like a modern e-commerce or library catalog. The result is a system that is easier for catalogers to apply and more intuitive for users, particularly in digital environments. This research differs from prior studies in several ways. While previous work utilized detailed input data and focused on structured outputs in MARC format, this research operates with limited data and centers on FAST subject headings. Additionally, this study uses a large, stratified

dataset, allowing for deeper analysis of AI output accuracy. Most importantly, this research addresses immediate challenges with practical implications while contributing to broader advancements in library cataloging.

Call Numbers and Titles for Subject Headings

Given this study's reliance on call numbers and titles for generating subject headings, it is useful to consider research examining the semantic relationships between title phrases and subject terms. This section also evaluates the role of call numbers in achieving accurate classification and explores the distinctions between structured and faceted subject headings. In contrast to the more contemporary literature on AI's role in cataloging and classification, this review draws primarily from earlier decades of research. Wise Fernandez (1991) studied title phrases and their semantic links to subject headings, mirroring our research goals. Their study evaluated LCSH's ability to represent book titles, revealing insights into subject terms' accuracy, precision, and vocabulary similarity. Only 26% of title phrases matched LCSH terms exactly, with 74% partial matches. Hierarchical mismatches were common, as titles were often more specific than LCSH terms. Wise Fernandez suggested expanding LCSH and adding title keyword searches to improve retrieval. Supporting Wise Fernandez, Engelson (2013) evaluated aligning title concepts with deconstructed LCSH terms. Engelson's study revealed inconsistencies in "fast-track cataloging" among catalogers. Match rates varied by discipline and format, requiring caution even in high-match areas. Engelson questioned adding LCSH subject headings when keyword searches drive retrieval.

The retrieval efficiency of LCSH has long been a topic of debate (Byrne, 1975). Alongside scholarly discussions, the Library of Congress reviewed the benefits and drawbacks of pre-coordinated (LCSH) versus post-coordinated (faceted vocabularies) subject cataloging, prompting changes to cataloging process (Cooey & Phillips, 2003). Post-coordination offers greater flexibility for catalogers in creating records and increased efficiency for users when searching for materials. Recently, the practice of assigning individual headings to resources, rather than constructing intricate subject strings, has gained popularity. A notable example is the British Library's adoption of FAST in early 2022 (Pyrne, 2023). This present study capitalizes on the simplicity of post-coordination to support cataloging efforts, enhancing brief bibliographic records for effective keyword searching. Faceted vocabularies provide a promising solution to address challenges in subject generation, particularly with minimal input data. Despite data constraints, FAST's compatibility with LCSH ensures effective retrieval without compromising cataloging.

Methods

A stratified sample of 1,000 records was selected from the 68,000 MARC records with LCCNs, allowing access to their corresponding full records from OCLC for detailed analysis and extrapolation to the broader population of 212,000 brief catalog entries. These records were batch-searched through OCLC Connexion to retrieve additional metadata, particularly FAST subject headings. To reflect the dataset's composition, the sample was stratified into five domains based on LC call number classifications: history (277 records), humanities (277), social sciences (243), arts (83), and sciences (120). The evaluation focused on FAST subjects, as its faceted, term-based structure proves more adaptable to AI simulation and generation. LCSH's pre-coordinated, structured format (e.g., subdivided headings) presents greater obstacles for

replication, especially given the limited context provided by brief titles. While both LCSH and FAST contribute to effective retrieval, FAST's straightforward design aligns more closely with the AI-driven methodology employed in this study.

A generative AI tool (Grok 3, xAI) was programmed to generate subject terms based on pre-trained FAST rules. Inputs were limited to LC call numbers and title/author data, with OCLC metadata reserved for comparison. The AI was configured to produce one to four subject terms per record, reflecting typical FAST usage patterns. A confidence score (0.0 to 1.0) was assigned by AI to each record, measuring term specificity and contextual clarity according to the AI's alignment with FAST vocabularies. Higher scores indicate stronger semantic alignment with the controlled vocabulary. Below is the finalized prompt used for Grok's output generation. It has been tested multiple times to ensure alignment with AI best practices. The prompt provides clear instructions for handling cases where no suitable FAST terms are identified, calculating confidence scores, and managing errors or unstructured data. Due to Grok's limitations in processing large datasets, this research processed 100 records per run until all 1,000 records were completed.

> **Prompt**: Using the uploaded Excel file containing the columns "LC Call Number" and "Title/Author," to assign 1–4 Faceted Application of Subject Terminology (FAST) subject terms per record, strictly adhering to FAST subject heading guidelines. Ensure the output is formatted for seamless integration into Excel, with four columns: "LC Call Number," "Title/Author," "Subject," and "Confidence Score." The "Subject" column should list all assigned terms for a record in a single cell, separated by semicolons. The Confidence Score, ranging from 0.0 to 1.0, will reflect the semantic relevance and adherence to FAST guidelines for each term assignment. If no suitable FAST terms can be assigned, populate the "Subject" column with None and set the Confidence Score to 0.0. Handle invalid or incomplete data by flagging records with a note in the "Subject" column (e.g., "Invalid Data") and a Confidence Score of 0.0.

The author compared AI-generated terms to OCLC FAST subjects in a two-step process. First, direct matching sequentially compared each AI term to the FAST subject list for exact or near-exact equivalence. The author calculated a match score as the number of matches divided by the number of AI terms (M/T); distributed scores can be found in Table 1.

TABLE 1
Match Score Distribution

Number of AI Terms	Number of Matches = Match Score			
1	1 = 1.0			
2	1 = 0.5	2 = 1.0		
3	1 = 0.33	2 = 0.67	3 = 1.0	
4	1 = 0.25	2 = 0.5	3 = 0.75	4 = 1.0

Given the complexity and variability of subject vocabularies, direct matching alone may not fully assess term relevance, as cataloging lacks strict standardization. Thus, a second step, enhanced matching, incorporated semantic mapping, thematic alignment, and conceptual relevance assessment, leveraging FAST's hierarchical and associative relationships. Enhanced matching drew on the AI's training across a vast text corpus, simulating subject matter expertise by referencing ontologies, taxonomies, and controlled vocabularies (e.g., FAST, LCSH). This external validation enabled the AI to identify valid terms beyond exact matches, such as

linking "Industrial Revolution" to "economic history." The author calculated and aggregated three performance metrics per record:

1. Coverage (Recall): The percentage of FAST terms captured by AI terms, reflecting completeness. Formula:

$$C = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{M_i}{T_i}\right) \times 100$$

where C = coverage, N = total records (1,000), M_i = matched terms in record i, T_i = total FAST terms in record i.

2. Precision: The percentage of AI terms that are valid (i.e., match or align with FAST terms), reflecting accuracy. Formula:

$$P = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{V_i}{T_i}\right) \times 100$$

where P = precision, N = total records (1,000), V_i = valid AI terms (direct or enhanced matches) in record i, Ti = total AI terms in record i.

3. F1 Score: The harmonic mean of coverage and precision, balancing both. Formula:

$$F1 = 2 \times \frac{P \times C}{P + C}$$

where P = precision, C = coverage (Recall).

Precision, recall, and F1 score are standard metrics in information retrieval. They evaluate system output against a gold standard. They balance accuracy and completeness in classification and metadata tasks (Manning et al., 2008).

After calculating coverage and precision, the author derived two correlation metrics. The direct correlation (r) is the Pearson correlation coefficient between the average confidence score and the coverage rate, reflecting the strength of the linear relationship between them. The enhanced correlation (r) is the Pearson correlation coefficient between the average confidence score and the precision rate, assessing the strength of the linear relationship. The author then aggregated match scores, coverage, precision, and F1 scores (mean, SD) across all records and by domain. Additionally, Pearson correlation coefficients (r) assessed relationships between (1) average confidence scores and direct match scores (direct correlation) and (2) average confidence scores and enhanced match scores (enhanced correlation), with p-values testing significance. Visualizations, including histograms of match scores, bar charts of coverage/precision/F1, and scatterplots of confidence versus match scores, supported the analysis.

Findings

Matching Metrics

AI-generated subject terms achieved an average direct match score of 0.85 (SD = 0.14) and an enhanced match score of 0.95 (SD = 0.10) across 1,000 records (see Table 2). Among the five domains, history exhibited the highest match scores (Direct: 0.88, SD = 0.13; Enhanced: 0.97, SD = 0.08), while sciences recorded the lowest (Direct: 0.80, SD = 0.17; Enhanced: 0.92, SD = 0.12). Humanities, social sciences, and arts fell within a similar range, with direct matches between 0.82 and 0.85 and enhanced matches between 0.94 and 0.96. Confidence scores averaged 0.87 (SD = 0.08). Direct matching yielded scores ≥0.75 for 71% of records (710/1,000), rising to 90%

(900/1,000) with enhanced matching. Variability in direct matching (SD = 0.14) suggests moderate dispersion, with sciences exhibiting the greatest variation (SD = 0.17) and history the lowest (SD = 0.13). Enhanced matching reduced variability across all domains, with SDs ranging from 0.08 to 0.12, reflecting improved alignment. The greater consistency in enhanced matching, particularly in history and social sciences, suggests stronger semantic refinement, making confidence scores a more reliable predictor of match quality. Improved term mapping, such as aligning "political culture" with "politics and government," will enhance retrieval accuracy. The overall direct match score of 0.85 indicates solid baseline accuracy, but enhanced matching (0.95) significantly improves alignment with FAST subjects. History (0.97) outperforms sciences (0.92) in enhanced matching, likely due to frequent exact term matches (e.g., "history") and strong thematic relationships (e.g., "social conditions" linking to "economic conditions"). In contrast, sciences' lower score may stem from specialized terminology (e.g., "tapeworms" versus "Bothriocephalidae"), which requires deeper taxonomic mapping.

TABLE 2
Match Scores by Domain

Domain	N	Direct Match (Mean)	Direct Match (SD)	Enhanced Match (Mean)	Enhanced Match (SD)
History	277	0.88	0.13	0.97	0.08
Humanities	277	0.82	0.16	0.94	0.11
Social Sciences	243	0.85	0.14	0.96	0.09
Arts	83	0.83	0.15	0.95	0.1
Sciences	120	0.8	0.17	0.92	0.12
Overall	**1,000**	**0.85**	**0.14**	**0.95**	**0.1**

Precision, Recall, and F1 Scores

Precision, recall, and F1 scores were computed based on 2,495 AI-generated terms and 2,499 FAST terms. Direct matching yielded precision (0.64), recall (0.50), and F1 score (0.56). Enhanced matching significantly improved these metrics, achieving precision (0.87), recall (0.68), and F1 score (0.76). Among the five domains, history led in F1 score (0.81), while sciences scored lowest (0.72) (see Table 3).

TABLE 3
Enhanced Matching Metrics by Domain

Domain	Precision	Recall	F1 Score
History	0.91	0.72	0.81
Humanities	0.85	0.66	0.74
Social Sciences	0.89	0.7	0.78
Arts	0.86	0.67	0.75
Sciences	0.83	0.64	0.72
Overall	**0.87**	**0.68**	**0.76**

As Table 4 shows, enhanced matching significantly outperformed direct matching, achieving higher precision (0.87 versus 0.64), recall (0.68 versus 0.50), and F1 score (0.76 versus 0.56).

Confidence Score Correlations

The histogram in Figure 1 reveals a bimodal distribution of direct match scores, peaking at 0.5 (240 records) and 1.0 (150 records), with a smaller tail (0.0–0.4, ~460 records below 0.75). Figure 2 demonstrates a right-skewed distribution for enhanced matching scores, peaking at 1.0 (700 records), with fewer outliers. The confidence score, which measures term specificity and contextual clarity in alignment with FAST vocabularies, serves as a reliable

TABLE 4
Precision, Recall, and F1 Scores

Metric	Direct Matching	Enhanced Matching
Precision	0.64	0.87
Recall	0.5	0.68
F1 Score	0.56	0.76

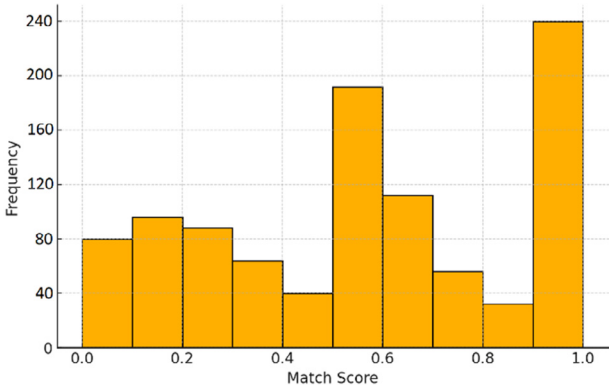


Figure 1. Histograms of Direct Match Scores Across 1,000 Records.

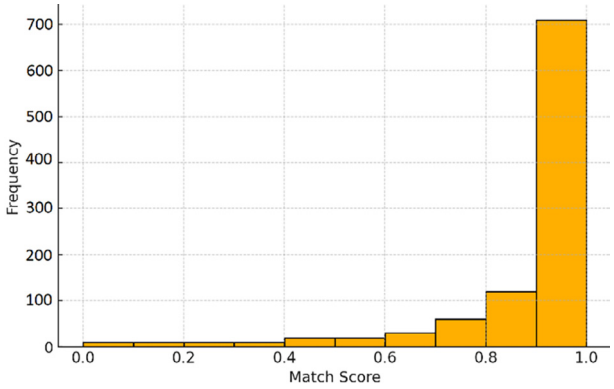


Figure 2. Histograms of Enhanced Match Scores Across 1,000 Records.

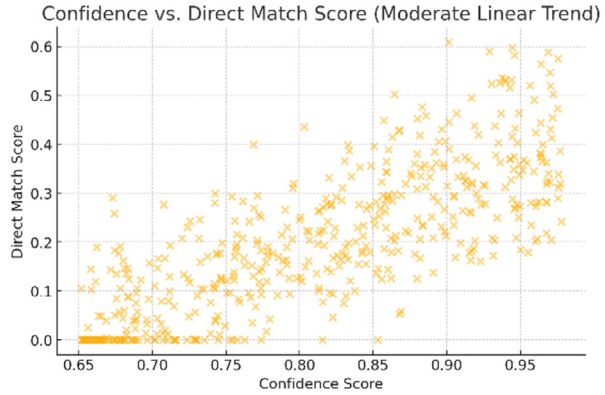


Figure 3. Scatterplot of Direct Match Scores Versus Confidence Scores.

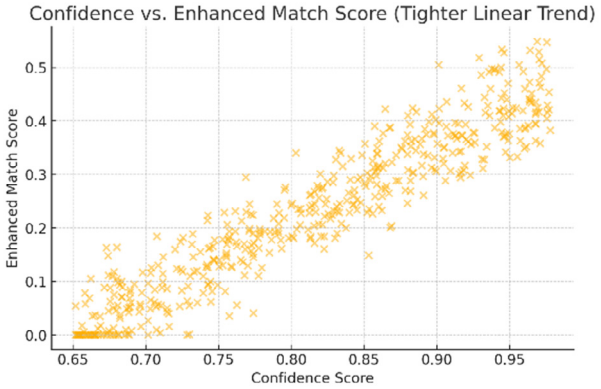


Figure 4. Scatterplot of Enhanced Match Scores Versus Confidence Scores.

indicator of conceptual relevance, as evidenced by the tighter clustering of enhanced match scores, suggesting that confidence better reflects semantic alignment between AI-generated terms and FAST headings.

Pearson's correlation coefficients (see Table 5) indicate a moderate correlation ($r = 0.62$, $p < 0.01$) between confidence scores and direct match scores, while the stronger correlation ($r = 0.78$, $p < 0.01$) with enhanced match scores suggests that confidence better reflects semantic alignment.

Figure 3 (Direct Matches) reveals a moderate relationship between confidence scores and match scores but with substantial spread. Figure 4 (Enhanced Matches) shows tighter clustering, reinforcing that enhanced matching better aligns confidence with actual relevance.

Examples

Table 6 presents five brief records, each accompanied by subject terms generated by the AI tool and by human catalogers as recorded in OCLC's MARC records using the FAST standard.

TABLE 5
Correlation Results

Comparison	Pearson *r*	*p*-Value	Interpretation
Confidence vs. Direct	0.62	<0.01	Moderate positive correlation
Confidence vs. Enhanced	0.78	<0.01	Strong positive correlation

TABLE 6
Examples of Brief Records with AI and FAST Subject Terms

Call Number	Title/Author	AI Subject	Confidence Score	FAST Subject	Direct Match	Enhanced Match		
					No #	Score	No #	Score
E169.1.W82 1955	*Culture on the moving frontier* / Louis B. Wright	Frontier and pioneer life; United States; Culture	0.95	Civilization; Frontier and pioneer life; United States	2	0.67	3	1.0
HD4903.5.U6U55 1969	*Jobs & civil rights: the role of the federal government in promoting equal opportunity in employment and training* / Richard P. Nathan	Discrimination in employment; United States; Civil rights; 1960-1970	0.95	Discrimination in employment; Law and legislation; Affirmative action programs; United States; Equal Employment Opportunity Commission; Civil Rights Act of 1964 (United States)	2	0.50	3	0.75
PN1042.H34	*The seven principles of poetry* / Anne Hamilton	Poetry; Poetics; Literary criticism	0.91	Literature–Philosophy; Poetry; Poetics	2	0.67	3	1.0
HC254.5.M625 1964	*England on the eve of the Industrial Revolution: A study of economic and social conditions from 1740 to 1760, with special references to Lancashire* / Louis W. Moffit	Industrial revolution–England; England–Economic conditions; 18th century; Lancashire (England)	0.97	Economic history; Social conditions; England–Lancashire	1	0.25	3	0.75
QP361.S9 1967	*Physiological and biochemical aspects of nervous integration* / Francis D. Carlson	Neurophysiology; Biochemistry; Nervous System	0.98	Neurochemistry; Nervous System; Electrophysiology	1	0.33	3	1.0

Terms in the table are separated by a semicolon. The examples illustrate the application of a relatively strict measure, where chronological terms like "1964" are not considered equivalent to a range such as "1960-1970," and enhanced matches are limited to clear hierarchical or contextual relationships.

For the first record, *Culture on the Moving Frontier*, the AI generated three subject terms: "frontier and pioneer life," "United States," and "culture," while the FAST headings included "frontier and pioneer life," "United States," and "civilization." Direct matching identified exact matches for the first two sets of terms, yielding a score of 0.67 (2/3). Enhanced matching recognized "culture" as a related concept to "civilization," despite lacking a direct hierarchical link in FAST, resulting in an enhanced match score of 1.0 (3/3). The second record has four AI-produced terms and six FAST subject terms. Direct matching confirmed exact matches for "discrimination in employment" and "United States," but "civil rights" and "1960-1970" did not match exactly, as "1964" is a specific year, not a range, resulting in a direct match score of 0.5 (2/4). Enhanced matching linked "civil rights" to "Civil Rights Act of 1964 (United States)" as a direct manifestation, but "1960-1970" was not matched due to the strict chronological rule, leading to an enhanced match score of 0.75 (3/4).

For *The Seven Principles of Poetry*, both AI and FAST have three terms. Direct matching identified exact matches for "poetry" and "poetics," with a direct match score of 0.67 (2/3). Enhanced matching treated "literary criticism" as a partial match to "literature–philosophy," as the concepts are related but not identical in FAST's vocabulary, resulting in an enhanced match score of 1.0 (3/3). The record for *England on the Eve of the Industrial Revolution* has four AI terms and four FAST terms. Direct matching identified near-exact matches for "Lancashire–England," giving a direct match score of 0.25 (1/4). Enhanced matching partially matched "Industrial Revolution–England" to "economic history" as a broader concept and "England–economic conditions" to "social conditions" as a related concept, while "18th century" remained unmatched, leading to an enhanced match score of 0.75 (3/4). For this record, the AI-generated terms include a chronological facet, "18th century," which the human-created FAST terms lack, highlighting a difference in cataloging approaches.

The final record has three AI terms and three FAST terms. Direct matching identified an exact match for "nervous system," giving a direct match score of 0.33 (1/3). Enhanced matching partially matched "neurophysiology" to "electrophysiology" (a subfield) and "biochemistry" to "neurochemistry" (a narrower term), leading to an enhanced match score of 1.0 (3/3). These examples highlight the variability in alignment between AI-generated terms and FAST headings across domains, with direct match scores ranging from 0.33 to 0.67 and enhanced match scores from 0.67 to 1.0. It reflects the stricter criteria and the challenges of specialized terminology in sciences (e.g., the last record in Table 6: *Physiological and Biochemical Aspects of Nervous Integration*).

Discussion

The strong performance of enhanced matching (F1 score: 0.76) supports the use of AI as a reliable tool for enriching brief MARC records. AI generates standardized subject headings accurately, even with minimal metadata. Comparatively, direct matching misses some valid FAST terms due to limited input data. It performs less effectively than enhanced matching, which uses semantic relationships. This limitation needs context, however. FAST terms used for comparison are created by human catalogers, and their choices of subject headings (and

even call numbers) vary based on judgment and knowledge (Chow et al., 2024). This variability happens even when they have access to the books being cataloged. As a result, expecting a perfect match rate between AI terms and FAST headings is not realistic or necessary.

As expected, AI tends to provide general terms. Its input data is limited to LC call numbers, titles, and authors, but author data helps little unless the authors are famous. This indicates that training sources for some AI tools are restricted. The broad nature of AI terms leads to imperfect direct matches between AI-generated terms and FAST headings. On the other hand, AI performs better with more specific book titles. For example, when a title includes a famous figure like Shakespeare, AI generates terms that align more closely with FAST headings. Domain-specific variability in match scores highlights AI's strengths and limitations. While history records benefit from broader thematic consistency (F1 score: 0.81), sciences require more nuanced taxonomic mapping (F1 score: 0.72). This aligns with Zhang et al.'s (2023) observation. They note that AI struggles with discipline-specific rules without sufficient contextual training. AI excels at generating comparable FAST terms for batch processing large collections, like the 212,000 brief records in this study. However, human review remains essential in the workflow. Human review should prioritize domains with lower recall or records with low confidence scores.

The moderate-to-strong correlation between confidence scores and match quality (r = 0.62–0.78) suggests that libraries can use these scores to triage records for review. For instance, records with confidence scores ≥0.9 (which comprised 62% of the sample) could be auto-approved, while scores ≤0.5 might flag records for human intervention. The best way of future fine-tuning may be collaborative projects in between catalogers and subject librarians. AI's reliance on titles and call numbers limited its ability to infer granular subjects and expect AI to have more domain-specific training to reduce false negatives. Future iterations could incorporate scientific taxonomies (e.g., medical subject headings for sciences) to improve recall, addressing the challenges observed in *Chest Disorders in Children* and *Physiological and Biochemical Aspects of Nervous Integration* (see Table 6). This study relies on data from a single institution's collection, which needs to be strengthened for the external validity by testing across diverse libraries, if possible. Moreover, a longitudinal study comparing circulation rates before and after AI metadata enhancement, particularly for history records like *Culture on the Moving Frontier*, could quantify the impact on user discovery.

Conclusion

This study demonstrates AI's potential to enhance cataloging for legacy collections, addressing the challenge of limited metadata in 212,000 brief MARC records at the author's library. By leveraging minimal data—such as call numbers, titles, and authors—AI effectively generates standardized subject terms, improving discoverability for users. Additionally, the integration of FAST vocabularies, with their simplified, faceted structure, supports this automated process and enhances cataloging efficiency. The findings of this study suggest promising implications for library cataloging. AI-driven automation has the capacity to substantially reduce human labor. A study on copy-cataloging productivity estimates an average rate of 1.4–1.7 records per hour (Buschman & Chickering, 2007). In theory, if a cataloger worked eight hours per day, five days per week, for 52 weeks per year, and processed 1.7 records per hour, they would complete approximately 3,536 records annually. To manually catalog 212,000 brief records without breaks and assuming just copy cataloging were needed, an estimated 60 staff

members would be required to work for a year. This underscores AI's potential to significantly streamline cataloging operations.

It is worth noting that AI tends to generate overly broad subject terms, particularly in scientific domains, necessitating human oversight to maintain accuracy. A hybrid approach—combining AI automation with expert review—presents a practical solution that balances efficiency with quality control. As libraries continue to adopt faceted vocabularies like FAST, AI-assisted cataloging offers a scalable method for improving access to under-described collections by augmenting brief MARC records with subject headings. Future research should focus on domain-specific AI training to refine accuracy and applicability. Additionally, testing across diverse library collections will further strengthen AI's role in cataloging, ensuring institutions worldwide can optimize discoverability and accessibility.

Acknowledgments

This research was conducted as part of my sabbatical work. I want to thank the library dean for approving my sabbatical request and my colleagues who took my job responsibilities during the leave time and supported me throughout the entire process.

References

Buschman, J., & Chickering, F. W. (2007, July). A rough measure of copy cataloging productivity in the academic library. *Library Philosophy and Practice*, (139). https://digitalcommons.unl.edu/libphilprac/139

Byrne, J. R. (1975). Relative effectiveness of titles, abstracts, and subject headings for machine retrieval from the COMPENDEX services. *Journal of the American Society for Information Science*, *26*(4), 223–229. https://doi.org/10.1002/asi.4630260405

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. *Cataloging & Classification Quarterly*, *62*(5), 574–588. https://doi.org/10.1080/01639374.2024.2394516

Cooey, N., & Phillips, A. (2023). Library of Congress Subject Headings: A post-coordinated future. *Cataloging & Classification Quarterly*, *61*(5–6), 491–505. https://doi.org/10.1080/01639374.2023.2193584

Engelson, L. (2013). Correlations between title keywords and LCSH terms and their implication for fast-track cataloging. *Cataloging & Classification Quarterly*, *51*(6), 697–727. https://doi.org/10.1080/01639374.2013.779351

Frank, E., & Paynter, G. W. (2004). Predicting Library of Congress classifications from Library of Congress subject headings. *Journal of the American Society for Information Science and Technology*, *55*(3), 214–227. https://doi.org/10.1002/asi.10360

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Martins, S. (2024, February 27). *Artificial intelligence-assisted classification of library resources: The case of Claude AI*. DigitalCommons@University of Nebraska - Lincoln. https://digitalcommons.unl.edu/libphilprac/8159

Mödden, E. (2022). Artificial intelligence, machine learning and bibliographic control. DDC short numbers: Towards machine-based classifying. *JLIS.it, 13*(1), 256–264. https://doi.org/10.4403/jlis.it-12775

Pyrne, N. C. (2023). FAST: A journey toward sustainability in subject indexing at the British Library. *Catalogue & Index*, (211), 32–35. https://doi.org/10.1080/01639374.2023.2238701

OCLC. (2025). *FAST: Subject terminology schema*. https://www.oclc.org/en/fast.html

Wise Fernandez, C. (1991). Semantic relationships between title phrases and LCSH. *Cataloging & Classification Quarterly*, *13*(1), 51–77. https://doi.org/10.1300/J104v13n01_04

Zhang, S., Wu, M., & Zhang, X. (2023, October 17). Utilising a large language model to annotate subject metadata: A case study in an Australian national research data catalogue. *arXiv*. https://arxiv.org/abs/2310.11318