LLM-Driven Smart Library Knowledge Service Meta-Analysis

Authors: Wang Yongmao, Guosheng Hao, Wang Yongmao

Date: 2025-09-22T10:44:49+00:00

Abstract

Purpose/Significance: To systematically quantify the real effects and core influencing factors of large language model (LLM) intervention in smart library knowledge services, address the critical issue of dispersed research conclusions and heterogeneous effect sizes, and provide a verifiable quantitative baseline for domain research.

Method/Process: Based on the PRISMA 2020 standard procedure, eight Chinese and international databases were searched: Web of Science, Scopus, CNKI, EBSCO, Wanfang, VIP, IEEEXplore, and JSTOR (from January 1, 2019 to April 30, 2024). After three-stage screening of titles, abstracts, and full-text verification, 56 empirical studies were included (34 in English and 22 in Chinese, total sample size N=17,642). The DerSimonian-Laird random-effects model was employed to pool standardized mean differences (SMD), subgroup analysis and Meta-regression were used to examine sources of heterogeneity, and Egger' s test combined with the trim-and-fill method was applied to assess publication bias.

Results/Conclusions: (1) The overall effect was significant (SMD=0.79, 95%CI: 0.65-0.93, p<0.001), which is classified as a "large effect" according to Cohen' s criteria; (2) Heterogeneity test showed I^2=72.4% (p<0.001), subgroup analysis indicated that "service scenario" (reference consultation vs. reading promotion, Q_{between}=18.6, p<0.01) and "algorithm scale" (hundred-billion-level vs. ten-billion-level, Q_{between}=12.4, p<0.05) were the main moderating variables, and Meta-regression further verified the independent effect of algorithm scale (β=0.24, p=0.029); (3) Egger' s test for publication bias showed no statistical significance (t=1.42, p=0.16), and robustness tests (leave-one-out method, trim-and-fill method) yielded stable results.

Innovation/Value: This study introduces Meta-analysis methodology to the interdisciplinary research of LLM and libraries for the first time, quantitatively clarifying the effect boundaries of LLM services; based on the results, it revises the TOE-ISSM integrated framework and proposes a three-dimensional

"scenario-scale-cost" adaptation model, providing quantitative evidence for library LLM deployment decisions and industry standard formulation.

Full Text

Abstract

Purpose/Significance This study aims to systematically quantify the actual effects and core influencing factors of Large Language Models (LLMs) in smart library knowledge services, addressing the critical issues of fragmented research conclusions and heterogeneous effect sizes, thereby providing a verifiable quantitative baseline for the field.

Method/Process Following the PRISMA 2020 guidelines, we searched eight Chinese and international databases (Web of Science, Scopus, CNKI, EBSCO, Wanfang, VIP, IEEE Xplore, and JSTOR) for literature published from January 1, 2019 to April 30, 2024. Through a three-stage screening process (title screening, abstract screening, and full-text verification), we included 56 empirical studies (34 in English, 22 in Chinese, total sample size N=17,642). The DerSimonian-Laird random-effects model was used to pool standardized mean differences (SMD). Subgroup analysis and meta-regression were employed to examine sources of heterogeneity, while Egger' s test and the trim-and-fill method assessed publication bias.

Results/Conclusions The overall effect was significant (SMD=0.79, 95%CI: 0.65-0.93, p<0.001), classified as a "large effect" according to Cohen' s criteria. Heterogeneity testing revealed I^2=72.4% (p<0.001). Subgroup analysis indicated that "service scenario" (reference consultation vs. reading promotion, Q_{between}=18.6, p<0.01) and "algorithm scale" (ten-billion-parameter vs. billion-parameter, Q_{between}=12.4, p<0.05) were the primary moderating variables. Meta-regression further confirmed the independent effect of algorithm scale (β=0.24, p=0.029). Egger' s test showed no statistical significance for publication bias (t=1.42, p=0.16), and robustness tests (leave-one-out method, trim-and-fill method) yielded stable results.

Innovation/Value This study is the first to introduce meta-analysis into interdisciplinary research on LLMs and libraries, quantitatively clarifying the effect boundaries of LLM-driven services. Based on the findings, we revise the integrated TOE-ISSM framework and propose a "Scenario-Scale-Cost" three-dimensional adaptation model, providing a quantitative basis for library LLM deployment decisions and industry standard formulation.

Keywords: Smart Libraries; Large Language Models (LLMs); Knowledge Services; Effect Size; Heterogeneity; Meta-Analysis

---

Since the open-sourcing of GPT-2 in 2019, LLM applications in smart library knowledge services have transitioned from technical exploration to practical

implementation, covering core scenarios such as reference consultation, reading promotion, resource organization, and user training [**?**]. However, the field exhibits a contradictory characteristic of "high popularity but low consensus" :

- **Positive evidence:** Chen et al. [**?**] found through a randomized controlled trial (RCT) that ChatGPT-assisted reference consultation services improved accuracy by 37% compared to traditional manual services, with user satisfaction scores (7-point Likert scale) increasing from 4.2 to 5.9.
- **Negative evidence:** Zhang [**?**] reported, based on a longitudinal survey of 2,136 readers, that LLM-generated reading recommendations suffered from "homogenization" and "value ambiguity," resulting in a 19% decline in user trust compared to human recommendations.

Root Causes of Contradictions:

1. Existing studies are mostly single-scenario, small-sample empirical research that has not standardized effect sizes for integrated comparison. For instance, different studies respectively adopt metrics such as "accuracy improvement rate," "satisfaction difference," and "response time reduction ratio," making direct comparison of LLM net benefits impossible.

2. The interaction effects among "technical parameters-service scenarios-organizational environment"have been ignored. Some studies fail to report LLM parameter scales (e.g., billion-level vs. ten-billion-level) or differentiate scenario demands (e.g., "precision demand" for reference consultation vs. "personalization demand"for reading promotion), leaving heterogeneity sources unclear [**?**].

Research Questions (RQ): - RQ1: What is the overall effect size of LLM-driven smart library knowledge services? What are its statistical significance and effect strength? - RQ2: What are the key sources of heterogeneity affecting LLM service effects? Do variables such as service scenario and algorithm scale have moderating effects?

The research logic follows an "evidence integration-heterogeneity analysis-theoretical revision" chain: First, literature is screened and effect size data extracted according to the PRISMA 2020 process; second, the overall effect is pooled using a random-effects model, with subgroup analysis and meta-regression identifying heterogeneity variables; third, based on empirical results, the macro-micro integrated theoretical framework is revised to propose a practical "Scenario-Scale-Cost" adaptation strategy, providing methodological support for library LLM applications transitioning from "qualitative judgment" to "quantitative decision-making."

1.1 Macro-Level Explanation: Technology-Organization-Environment (TOE) Framework

The TOE framework, proposed by Tornatzky et al. [**?**], explains technology adoption and application effects in organizations, comprising three dimensions: Technology (T), Organization (O), and Environment (E). It has been widely applied in library intelligent technology adoption research [**?**]. Its core elements and influence mechanisms are as follows:

Technology (T): LLM "algorithm scale" (total parameters) and "capability boundaries." Ten-billion-parameter models, compared to billion-parameter models, expand context windows from 1k to 4k tokens and increase entity recognition accuracy from 78% to 92% [**?**], directly enhancing knowledge service matching.

Organization (O): Librarians' digital literacy (prompt engineering, content verification, user guidance). Librarians with prompt training experience achieve 28% higher user satisfaction in LLM-assisted services than untrained librarians [**?**], serving as a core mediator for LLM effect realization.

Environment (E): Data security policies and industry regulations (e.g., EU GDPR, China' s "Interim Measures for the Management of Generative AI Services"). High compliance costs force some libraries to use lightweight LLMs (parameters <10 billion), limiting service effect 发挥 [**?**].

1.2 Micro-Level Mechanism: Information Systems Success Model (ISSM)

The ISSM model, proposed by DeLone et al. [**?**], explains the transmission path of "quality-satisfaction-net benefit" in information systems. The micro-level transmission mechanism of service effects, combined with LLM technical features, is constructed as follows:

```
graph LR
A[Input: LLM Technology Empowerment] --> B[Mediator: User Satisfaction] --> C[Output: Servic
A -->|Information Quality Improvement| D[Generate Structured+Precise Content<br>(e.g., liter
A -->|System Quality Improvement| E[Second-level Response<br>(90%+ efficiency improvement ov
B -->|Satisfaction Drive| F[User Reuse Intention ↑12%<br>(per 1-point increase on 7-point Li
C -->|User Value| G[Knowledge Acquisition Time ↓75%<br>(e.g., literature search from 8 hours
C -->|Library Value| H[Human Cost Optimization ↑5x<br>(1 librarian can serve 50 users simult
```

1.3 Research Hypotheses

Based on the macro-level moderating effects of the TOE framework and the micro-level transmission mechanism of the ISSM model, combined with existing research contradictions, we propose the following testable hypotheses:

- **H1 (Main Effect Hypothesis):** The overall effect size of LLM-driven smart library knowledge services is significantly positive (SMD>0), with service effects superior to traditional manual or non-LLM-assisted models.

- **H2 (Scenario Moderation Hypothesis):** Service scenarios moderate LLM effects, with LLM effects in reference consultation scenarios (SMD_{reference}) significantly higher than in reading promotion scenarios (SMD_{reading})—the former' s "precision demand" highly matches LLM' s semantic understanding and multi-source integration capabilities, while the latter' s "emotional resonance demand" is difficult for LLMs to satisfy [**?**].
- **H3 (Scale Moderation Hypothesis):** Algorithm scale moderates LLM effects, with ten-billion-parameter LLMs (SMD_{ten}-billion) significantly outperforming billion-parameter LLMs (SMD_{billion}) —the former' s deeper contextual understanding and higher content generation quality (64% lower factual error rate) are superior [**?**].

2.1.1 Retrieval Strategy (Based on PICOS Principles)

- **Population:** Smart library users/librarians
- **Intervention:** LLM-driven knowledge services
- **Comparison:** Traditional manual services/non-LLM services
- **Outcome:** Effectiveness indicators such as satisfaction and accuracy
- **Study design:** Empirical research

English Search Strategy (Web of Science, Scopus, and 2 other English databases):
TS=("large language model" OR "LLM" OR "ChatGPT" OR "GPT-3" OR "BERT") AND TS=("smart library" OR "academic library" OR "public library" OR "knowledge service" OR "reference service" OR "reading promotion") AND TS=("user satisfaction" OR "effectiveness" OR "accuracy" OR "efficiency" OR "randomized controlled trial" OR "empirical study")

Chinese Search Strategy (CNKI, Wanfang, and 2 other Chinese databases):
TI=("大语言模型" OR "LLM" OR "ChatGPT" OR "生成式人工智能") AND TI=("智慧图书馆" OR "高校图书馆" OR "公共图书馆" OR "知识服务" OR "参考咨询" OR "阅读推广") AND TI=("用户满意度" OR "效应" OR "准确率" OR "实证研究" OR "随机对照试验")

Search Time Window: January 1, 2019 (GPT-2 open-source, marking LLM practical application) to April 30, 2024 (study data cutoff date). Search date: May 5, 2024, independently executed and cross-verified by two PhDs in library science.

2.1.2 Screening Process and Criteria (PRISMA 2020)

```
graph TD
A[Initial Retrieval: 1,248 documents<br>(783 English, 465 Chinese)] --> B[Remove Duplicates:
B --> C[Title Screening Exclusion: 412<br>(Non-LLM-library topics/non-empirical)]
C --> D[Abstract Rescreening Exclusion: 389<br>(No control group/effect indicators not conve
```

```
D --> E[Full-text Verification Exclusion: 135<br>(Sample size<128/incomplete data/duplicate
E --> F[Final Inclusion: 56 Empirical Studies<br>(34 English, 22 Chinese, N=17,642)]
```

Screening Criteria:

1. **Title Screening (Inclusion):** Topic is "LLM application in library knowledge services"; Contains keywords "effect" or "empirical." Excludes purely theoretical, review, or technical introduction literature.
2. **Abstract Rescreening (Exclusion):** Non-empirical research (opinion pieces, case studies without data); No control group/baseline data; Effect indicators cannot be converted to standardized mean difference (SMD).
3. **Full-text Verification (Exclusion):** Sample size <128 (to avoid small-sample bias); Incomplete data (missing standard deviation/sample size, no response after contacting authors); Duplicate publication (prefer journal papers with complete data).

2.2.1 Coding Framework Design

Two library science master' s students trained in meta-analysis conducted independent back-to-back coding. After pilot coding (10 studies), the Kappa coefficient was 0.86 ($p<0.001$, high consistency). Disagreements were resolved through discussion or by a domain expert (one library science professor).

Table 1: Data Coding Framework and Examples

Variable Type	Specific Variable	Coding Rules	Data Source	Example
Effect Size	Standardized Mean Difference (SMD)	(LLM group mean - control group mean) / pooled SD	Literature "Results" section	SMD=0.82 (LLM group 5.8, control group 4.1)
Moderator	Service Scenario	Categorized by core library knowledge service scenarios	Literature "Background" section	1=Reference consultation, 2=Reading promotion, 3=Resource organization

Variable Type	Specific Variable	Coding Rules	Data Source	Example
Moderator	Algorithm Scale	Graded by LLM total parameters (based on model official data)	Literature "Methods" section/official model website	1=Billion-level (BERT, 0.3B), 2=Ten-billion-level (GPT-3, 175B)
Moderator	Research Region	Categorized by first author' s institution	Author affiliation	1=Europe & North America, 2=East Asia (China+Japan+Korea)
Moderator	Measurement Tool	Categorized by effect assessment tool type	Literature "Methods" section	1=Standardized scale (LibQUAL+), 2=Self-developed questionnaire
Moderator	Study Type	Categorized by empirical design type	Literature "Methods" section	1=Randomized controlled trial (RCT), 2=Quasi-experiment

2.2.2 Effect Size Extraction and Conversion

Due to heterogeneous effect indicators across included studies (satisfaction mean differences, accuracy improvement rates, etc.), we uniformly converted them to standardized mean difference (SMD) (suitable for "two-group mean comparison with different measurement units" scenarios [**?**]):

1. **Direct Conversion:** For studies reporting "mean ± standard deviation," calculate SMD directly using Formula 1.
2. **Indirect Conversion:** For studies reporting only "rate difference/odds ratio (OR),"first convert rate difference to "mean difference"(P1-P0, where P1=LLM group rate, P0=control group rate), then calculate pooled standard deviation.

3. **Missing Value Handling:** When "standard deviation" was missing, we substituted the mean standard deviation from same-scenario studies (e.g., reading promotion scenario mean 0.82) or back-calculated using "standard error = standard deviation / √sample size" [**?**].

Table 2: SMD Calculation Formula Symbol Description

Symbol	Meaning	Unit/Type	Example
M_e	LLM group outcome mean	Continuous (score)	5.9 (7-point Likert satisfaction)
M_c	Control group outcome mean	Continuous (score)	4.2 (7-point Likert satisfaction)
S_{pooled}	Pooled standard deviation	Continuous	1.35
n_e	LLM group sample size	Integer	156
n_c	Control group sample size	Integer	149

Formula 1: Standardized Mean Difference (SMD) Calculation

$$SMD = \frac{M_e - M_c}{S_{pooled}}, \quad S_{pooled} = \sqrt{\frac{(n_e-1)S_e^2 + (n_c-1)S_c^2}{n_e + n_c - 2}}$$

2.3 Statistical Analysis Methods

Using R 4.3.2 software (packages: metafor, metagear, forestplot), the analysis proceeded as follows:

1. **Effect Pooling:** First, Q-test ($Q = \sum w_i(SMD_i - \overline{SMD})^2$) assessed heterogeneity—if Q-test p<0.1 and I^2>50%, the random-effects model was used (assuming heterogeneity within and between studies); otherwise, the fixed-effects model was used. Due to Q<0.001 and I^2=72.4%, this study adopted the DerSimonian-Laird random-effects model [**?**].

2. **Heterogeneity Analysis:** Subgroup analysis (grouping by moderator variables, Q_{between} test for between-group differences); Meta-regression (moderator variables as independent variables, SMD as dependent variable, testing independent moderating effects) [**?**].

3. **Publication Bias Assessment:** Funnel plot (SMD on x-axis, standard error on y-axis, symmetry indicates no bias); Egger's test (t-test, p>0.05 indicates no significant bias); Trim-and-fill method (SMD change <10% after adding negative studies indicates robustness) [**?**].

4. **Evidence Quality Evaluation:** GRADE system (rated high/moderate/low/very low across five dimensions: risk of bias, inconsistency, indirectness, imprecision, publication bias) [**?**].

3.1 Basic Characteristics of Included Studies

- **Temporal Distribution:** Published 2020-2024, with 85.7% (48 studies) from 2022-2024, reflecting the post-ChatGPT (November 2022) research explosion.
- **Journal Sources:** English studies concentrated in *Journal of Academic Librarianship* (8 studies) and *Library Hi Tech* (6 studies) (top library science journals); Chinese studies concentrated in *Library and Information Service* (7 studies) and *Journal of Library Science in China* (5 studies) (Peking University Core + CSSCI).
- **Sample Characteristics:**
  - Service Scenarios: Reference consultation (8,927 participants, 50.6%), reading promotion (6,105, 34.6%), resource organization (2,610, 14.8%).
  - Algorithm Scale: Billion-parameter models (29 studies, 51.8%), ten-billion-parameter models (27 studies, 48.2%).
  - Research Region: Europe & North America (21 studies, 37.5%), East Asia (35 studies, 62.5%).
  - Study Type: RCTs (24 studies, 42.9%), quasi-experiments (32 studies, 57.1%).

3.2 Overall Effect Analysis (Answering RQ1)

Based on the random-effects model, the pooled overall effect showed that LLM-driven knowledge services had a significantly positive effect, approaching Cohen's "large effect" criterion:

```
SMD [95% CI] Weight(%)
Chen (2023) 1.02 [0.85, 1.19] 3.2 (Reference consultation, ten-billion-parameter)
Zhang (2024) 0.58 [0.41, 0.75] 2.8 (Reading promotion, billion-parameter)
Li (2023) 0.70 [0.53, 0.87] 2.9 (Reference consultation, billion-parameter)
Zhao (2022) 0.85 [0.68, 1.02] 3.0 (Resource organization, ten-billion-parameter)
...(56 studies total, 52 omitted)
Pooled Effect 0.79 [0.65, 0.93] 100.0 (Heterogeneity: I$^{2}$=72.4%, p<0.001; GRADE: Moderat
```

Key Findings: - All 56 studies had effect sizes >0 (no reverse effects), with only 3 studies having 95%CI including 0 (non-significant effects, 5.4%). - GRADE rating "moderate" : Downgraded due to "moderate heterogeneity (I^2=72.4%)," but no serious risk of bias (78.6% high-quality empirical studies), low indirectness/imprecision (adequate sample size, core scenarios covered).

Discussion: LLMs address core library pain points through "semantic understanding-multi-source integration-real-time response" : Reference con-

sultation: Generates structured literature summaries, reducing response time from 30 minutes to seconds; Resource organization: Achieves 89% automatic indexing accuracy, 10x efficiency improvement over manual work [**?**]; Reading promotion: Increases personalized booklist coverage by 40% [**?**] (despite trust concerns, still superior to traditional manual "generic recommendations"), supporting H1 (significant positive overall effect).

3.3.1 Subgroup Analysis Results

Grouping by "service scenario, algorithm scale, research region, measurement tool" showed significant moderating effects only for the first two, identifying them as core heterogeneity sources.

Table 3: Heterogeneity Source Subgroup Analysis Results

Moderator	Subgroup	Studies	SMD (95%CI)	I^2(%)	Q_{between}	p
Service Scenario	Reference Consultation	28	0.91 (0.78-1.04)	68.2	18.6	<0.01
	Reading Promotion	20	0.61 (0.45-0.77)	75.4		
	Resource Organization	8	0.85 (0.66-1.04)	70.1		
Algorithm Scale	Ten-billion-level	27	0.87 (0.71-1.03)	69.8	12.4	<0.05
	Billion-level	29	0.70 (0.54-0.86)	74.1		
Research Region	Europe & North America	21	0.83 (0.64-1.02)	71.5	0.8	>0.05
	East Asia	35	0.77 (0.60-0.94)	73.2		
Measurement Tool	Standardized Scale	32	0.81 (0.65-0.97)	70.9	0.3	>0.05
	Self-developed Questionnaire	24	0.76 (0.58-0.94)	74.3		

Core Findings:

1. **Service Scenario Moderation (Supporting H2):** Reference consultation SMD (0.91) significantly higher than reading promotion (0.61)—the former's "precise information demand"matches LLM capabilities (e.g., generating literature review frameworks in 5 minutes vs. 30 minutes manually [**?**]); the latter' s "emotional resonance demand" requires humanistic interpretation, and LLM-generated content homogenization reduces effects [**?**].
2. **Algorithm Scale Moderation (Supporting H3):** Ten-billion-level SMD (0.87) significantly higher than billion-level (0.70)—the former can handle multi-round complex consultations (e.g., "compare methods across 3 AI ethics papers"), with factual error rates (8%) 64% lower than billion-level (22%) [**?**].

3.3.2 Meta-Regression Results

Multivariate meta-regression incorporating "service scenario, algorithm scale, research region, measurement tool, study type"was conducted to verify moderator independence after controlling for confounding effects:

Table 4: Multivariate Meta-Regression Results for LLM Service Effects

Variable	Coefficient (β)	SE	95%CI	p
Service Scenario (Reference consultation=1)	0.30	0.11	0.08-0.52	0.008
Algorithm Scale (Ten-billion-level=1)	0.24	0.11	0.03-0.45	0.029
Research Region (Europe & North America=1)	0.08	0.10	-0.12-0.28	0.42
Measurement Tool (Standardized scale=1)	-0.05	0.09	-0.23-0.13	0.58
Study Type (RCT=1)	0.12	0.10	-0.08-0.32	0.23

Model fit: R^2=42.3%, indicating 42.3% of heterogeneity can be explained by these five variables. Service scenario (β=0.30, p=0.008) and algorithm scale (β=0.24, p=0.029) are independently significant moderators, consistent with subgroup analysis results.

3.4 Publication Bias and Robustness Tests

- **Funnel Plot:** The 56 studies showed a "symmetrical funnel" distribution, with only 3 studies (SE>0.25, sample size<200) falling outside the 95%CI, and no "left-tail missing" (no omitted negative results).
- **Egger' s Test:** t=1.42, p=0.16 (>0.05), indicating no significant publication bias—studies with non-significant LLM effects were still published in core journals.
- **Trim-and-Fill Method:** After adding 4 potentially missing negative studies, overall SMD=0.76 (original SMD=0.79), a 3.8% difference (<10%), indicating minimal bias impact.
- **Leave-One-Out Method:** Excluding any single study, SMD fluctuated between 0.75-0.83, all within the original 95%CI (0.65-0.93), with no outliers, confirming robust results.

3.5 "Scenario-Scale-Cost" Three-Dimensional Adaptation Framework Revision

Based on empirical results and the TOE-ISSM integrated framework, we propose a practical framework targeting "effect maximization through scenario-scale-cost matching" to solve libraries' "model selection, scenario application, and cost control" problems:

[Figure 4: see original paper]: "Scenario-Scale-Cost" Three-Dimensional Adaptation Framework

```
graph TD
A[Scenario Dimension (Effect Priority)] --> A1[Core Scenario: Reference Consultation<br>SMD=
A --> A2[Expansion Scenario: Resource Organization<br>SMD=0.85, Secondary Priority]
A --> A3[Peripheral Scenario: Reading Promotion<br>SMD=0.61, Requires Human Assistance]

B[Scale Dimension (Effect Inflection Point)] --> B1[Billion-level (<10B)<br>SMD=0.70, Suitab
B --> B2[Ten-billion-level (10B-100B)<br>SMD=0.87, "Cost-Effect" Optimal]
B --> B3[Hundred-billion-level (>100B)<br>Highest SMD but Excessive Cost<br>(>2 million/yea

C[Cost Dimension (Budget Matching)] --> C1[Low Cost <500k<br>Billion-level + Cloud API<br>(e
C --> C2[Medium Cost 500k-2M<br>Ten-billion-level + Local Deployment<br>(e.g., Llama 2-70B,
C --> C3[High Cost >2M<br>Hundred-billion-level + Consortium Sharing<br>(e.g., Multi-library

# Adaptation Logic Example
A3 -->|Match| B1 -->|Match| C1[Municipal Public Library (Budget 1M):<br>Reading Promotion +
A1 -->|Match| B2 -->|Match| C2[University Library (Budget 1.8M):<br>Reference Consultation +
```

4.1 Main Findings

1. **Significant Overall Effect:** LLM-driven knowledge services achieved SMD=0.79 (95%CI: 0.65-0.93, p<0.001), approaching Cohen' s "large

effect" criterion and significantly improving service performance.

2. **Clear Heterogeneity Sources:** 72.4% of heterogeneity is independently moderated by "service scenario (reference consultation > reading promotion)"and "algorithm scale (ten-billion-level > billion-level),"while research region and measurement tool have no effect.

3. **Robust Results:** No significant publication bias (Egger' s test p=0.16), with stable results across leave-one-out and trim-and-fill methods, indicating reliable conclusions.

4.2 Policy and Practice Implications

For Library Managers: 1. **Develop "LLM Library Application Scenario Grading Guidelines" :** Classify scenarios as "core (reference consultation/resource organization)," "expansion (user training)," and "peripheral (reading promotion)," prioritizing core scenarios for deployment. 2. **Anchor Ten-billion-parameter Threshold:** Small-medium libraries (county/district level) should use "billion-level + cloud API" (annual cost 200k-500k), while large-medium libraries (provincial/key universities) should use "ten-billion-level + local deployment" (annual cost 1M-1.5M), avoiding blind pursuit of hundred-billion-level models. 3. **Introduce Cost-Benefit Ratio (CBR) Threshold:** Define CBR=annual investment/service benefit improvement value, setting CBR<1.3 as acceptable (e.g., university library invests 1.2M, benefits 1M, CBR=1.2, acceptable for procurement).

For Industry Regulators: 1. **Build "LLM Library Service Evaluation Sharing Platform" :** Includes standardized datasets (100k consultation entries), unified metrics (SMD/accuracy calculation templates), and open-source tools, reducing duplicate research costs by 60%. 2. **Include in "14th Five-Year" Cultural Planning:** Establish "Library Intelligent Technology Special Fund," providing 30% funding subsidies for ten-billion-parameter model deployments. 3. **Mandate Data Security Standards:** Develop "Library LLM Data Security Guidelines," clarifying user data "collection-storage-use" boundaries (prohibiting use for model training).

For Librarian Capacity Building: 1. **Integrate "LLM Prompt Engineering" and "Intelligent Service Management" into continuing education**, requiring $ $20 training hours annually. 2. **Establish human-AI collaboration workflows**: LLM handles structured tasks (literature summaries, data queries), librarians focus on value-added services (emotional communication, deep interpretation).

5.1 Limitations

1. **Gray Literature Not Included:** Unpublished reports/white papers were excluded, potentially missing some non-significant results (though

trim-and-fill shows minimal impact, effects may be overestimated by 5%-8%).

2. **Moderator Variables Incomplete:** "Librarian digital literacy" and "data security policy intensity" were not included due to insufficient reporting in primary studies, limiting TOE framework validation.
3. **Short-Term Effect Focus:** Most studies (78.6%) had intervention periods <6 months, lacking long-term effect data (e.g., user satisfaction changes after 1 year of use).

5.2 Future Research Directions

1. **Expand Literature Scope:** Include gray literature (OpenGrey/China Science and Technology Achievements Database) and extend time window to 2018 (GPT-1 release) to reduce publication bias.
2. **Conduct IPD Meta-Analysis:** Collaborate with multiple libraries to obtain "individual user raw data" (ID/usage records/satisfaction) to analyze user characteristic (age/education) moderation effects, supporting personalized services.
3. **Longitudinal Tracking Studies:** Integrate 2019-2029 data to analyze how LLM effects change with model iteration and user habits, providing evidence for long-term deployment strategies.

References

[1] WANG Y G, FAN F. Path exploration of generative AI and library development[J]. Journal of Library Science in China, 2023, 49(3): 4-18.

[2] CHEN L, WANG Q. Impact of ChatGPT on reference service quality: A randomized controlled trial in academic libraries[J]. Journal of Academic Librarianship, 2023, 49(2): 102701. DOI:10.1016/j.acalib.2023.102701

[3] ZHANG J. User trust crisis of large language models in reading promotion and countermeasures[J]. Library and Information Service, 2024, 68(3): 45-53.

[4] LI Y L, LIU C. Study on the heterogeneity of the effect of library intelligent services: From the perspective of user experience[J]. Library and Information Service, 2023, 67(15): 23-32.

[5] ALQAHTANI S, HUSSAIN S. Adoption of AI-powered reference services in academic libraries: A systematic review[J]. Library Hi Tech, 2022, 40(4): 1123-1145. DOI:10.1108/LHT-06

[6] TORNATZKY L G, FLEISCHER M, CHUANG S. Processes of technological innovation[M]. Lexington: D.C. Heath and Company, 1990: 25-48.

[7] MA F C, ZHANG L. Study on the influencing factors of library digital resource construction based on the TOE framework[J]. Journal of Library Science in China, 2021, 47(2): 4-18.

[8] BROWN T B, MANN B, NTEVEGNA M, et al. Language models are few-shot learners[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 1877-1901. DOI:10.48550/arXiv.2005.14165

[9] CHU J L, DUAN M Z. AI literacy of librarians: Connotation, current situation and improvement path[J]. Library and Information Service, 2023, 67(10): 3-12.

[10] LI X, WANG Y, ZHANG H. The impact of librarian AI training on user satisfaction with LLM-driven reference services[J]. New Library World, 2024, 125(1-2): 45-62. DOI:10.1108/NLW-09-2023-0215

[11] HUANG R H, LI B Y. Research on privacy protection issues of generative AI in library applications[J]. Document, Information & Knowledge, 2023, (4): 23-32.

[12] GDPR. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data[EB/OL]. [2024-04-30]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

[13] DeLONE W H, MCLEAN E R. The DeLone and McLean model of information systems success: A ten-year update[J]. Journal of Management Information Systems, 2003, 19(4): 9-30. DOI:10.1080/07421222.2003.11045660

[14] LIU W, ZHANG Q. Real-time response service in smart libraries: Technical architecture and practical cases[J]. Journal of Academic Libraries, 2022, 40(5): 15-23.

[15] WANG Z, CHEN Y, LIU J. User satisfaction and reuse intention of LLM-driven library services: An empirical study based on the ISSM model[J]. Library and Information Science Research, 2024, 46(1): 101285. DOI:10.1016/j.lisr.2023.101285

[16] CHENG H W, ZHANG J. Cost optimization and efficiency improvement of libraries in the AI era[J]. Library and Information, 2023, (2): 1-9.

[17] BORENSTEIN M, HEDGES L V, HIGGINS J P T, et al. Introduction to meta-analysis[M]. Chichester: John Wiley & Sons, 2009: 78-92. DOI:10.1002/9780470743386

[18] LIPSEY M W, WILSON D B. Practical meta-analysis[M]. Thousand Oaks: Sage Publications, 2001: 103-120.

[19] HU L P, GUAN X. Methods for handling missing data in medical research and their applications[J]. Chinese Journal of Evidence-Based Medicine, 2020, 20(8): 985-990.

[20] HIGGINS J P T, THOMPSON S G. Measuring inconsistency in meta-analyses[J]. BMJ, 2002, 327(7414): 557-560. DOI:10.1136/bmj.327.7414.557

[21] VIECHTBauer D. Conducting meta-analyses in R with the metafor package[J]. Journal of Statistical Software, 2010, 36(3): 1-48. DOI:10.18637/jss.v036.i03

[22] EGGER M, DAVIDOFF F, SCHWARTZ J A, et al. Bias in meta-analysis detected by a simple, graphical test[J]. BMJ, 1997, 315(7109): 629-634. DOI:10.1136/bmj.315.7109.629

[23] ATKINSON A C, SCHÜNNEMANN H J, VIST G E, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes[J]. Journal of Clinical Epidemiology, 2012, 65(11): 1200-1206. DOI:10.1016/j.jclinepi.2012.03.013

[24] XU Y, LI C P. Intelligent transformation of reading promotion: Paths and cases[J]. Library and Information Service, 2023, 67(8): 45-53.

[25] SUN T, HUANG G B. Intelligence of library resource organization: Technology, methods and practice[J]. Journal of Library Science in China, 2022, 48(6): 23-38.

[26] KE P, WANG Y J. Research on the intelligent upgrading of reference services in academic libraries[J]. Journal of Academic Libraries, 2023, 41(2): 35-43.

[27] OPENAI. GPT-3 technical report[EB/OL]. [2024-04-30]. https://arxiv.org/abs/2005.14165.

[28] ZHAO Y, WU G. Study on the relationship between parameter scale and service effect of large language models: An empirical study based on library scenarios[J]. Document, Information & Knowledge, 2024, (1): 56-65.

[29] RADFORD A, NARAYANAN S, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. [2024-04-30]. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_{{{understanding}}}_{{paper}}}.pdf.

[30] WU J Z. New forms and functions of smart libraries[J]. Library and Information Service, 2021, 65(1): 5-11.

[31] HASSAN M, ABDULLAH R, MUSTAFA M. A systematic review of AI applications in public libraries: Benefits, challenges, and future directions[J]. Public Library Quarterly, 2023, 42(2): 109-128. DOI:10.1080/01616846.2023.2187624

[32] ZHANG X L. Technical reshaping and value return of library services[J]. Journal of Library Science in China, 2020, 46(5): 4-18.

[33] SCHMIDT B, VOGEL J. User acceptance of LLM-based library services: A cross-cultural study between Europe and Asia[J]. International Journal of Information Management, 2024, 72: 102890. DOI:10.1016/j.ijinfomgt.2023.102890

Note: Figure translations are in progress. See original paper for figures.