Navigating the AI-Driven Metadata Landscape: A Human Centered Approach

Ying-Hsang Liu, Ph.D.

Chemnitz University of Technology, Germany

Symposium: Metadata Standards Matter: Building a Sustainable Future for Information
19 March 2025

DublinCore

https://tinyurl.com/4b3s24dd

...nd clear survey   Language: English

Designed by Task Group of Metadata and AI

- Francisco Carlos Paletta (University of São Paulo) - Chair
- **Ying-Hsang Liu** (Chemnitz University of Technology, Germany) - Co-Chair

Translated and distributed by DCMI Education Committee Members

0%

Language: English   Change the language

Metadata and AI Survey of DCMI Education Committee

The **Survey on Metadata and AI** designed by the **DCMI Education Committee** intends to gather expert insights on the potential impact of AI on metadata creation and management within libraries and information services. You will be asked to rate your agreement with statements about the future role of AI tools, including generative and predictive AI. The survey explores AI's advantages, challenges, and ethical considerations, along with the essential skills librarians will need in an AI-powered environment. It addresses AI's influence on tasks like subject indexing, enhancing metadata quality, and linking data to external resources.

Thank you for participating in this study. Your expert opinion is valuable in shaping the future of AI applications in libraries and information services. Please respond to the following statements based on your knowledge and experience. The survey should take about 15 minutes to complete.

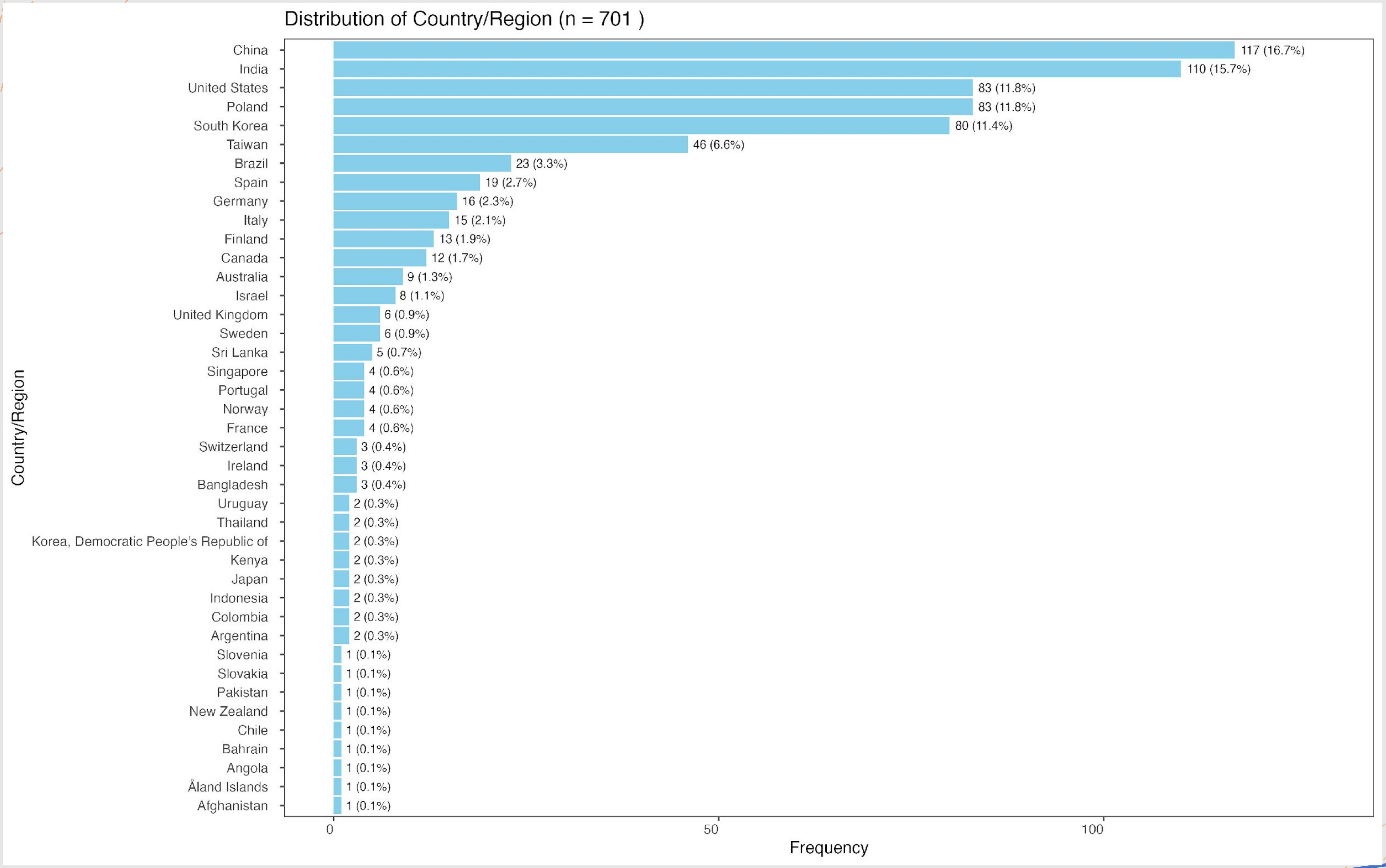
Next

Available in 14 languages

- हिन्दी
- தமிழ்
- 简体中文
- 繁體中文（台灣）
- 한국어
- 日本語
- Deutsch
- ✓ English
- Español
- Français
- Polski
- Português
- Português do Brasil
- Suomi

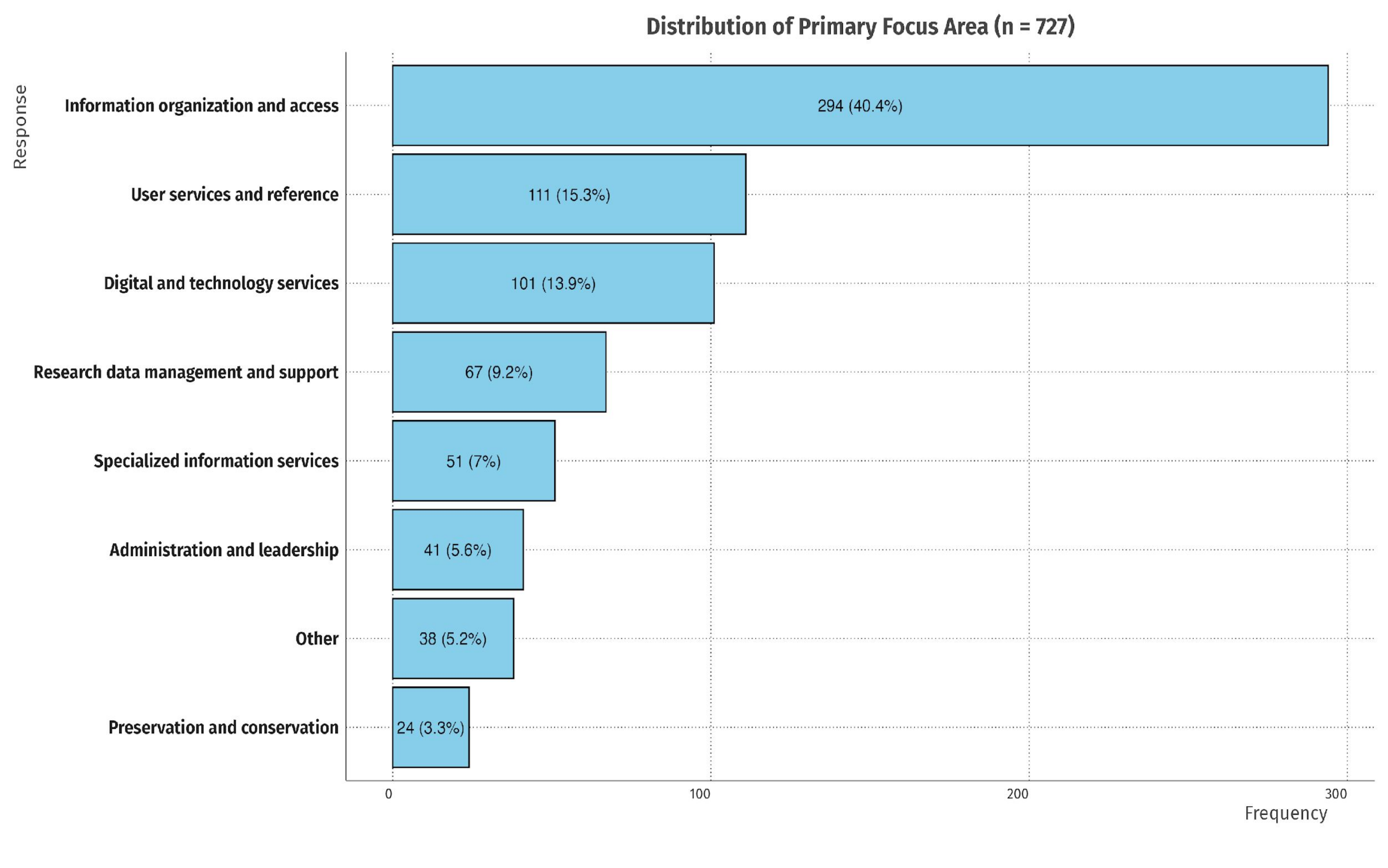
- Question items derived from literature, committee consultation and expert reviewing
- Four question groups:
  - **A.** Metadata Tasks and AI Applications
  - **B.** Potential Benefits, Challenges, and Concerns of Implementing AI in Metadata
  - **C.** Future Impact of AI on Metadata Creation and Management
  - **D.** Respondent Characteristics

Distribution of Country



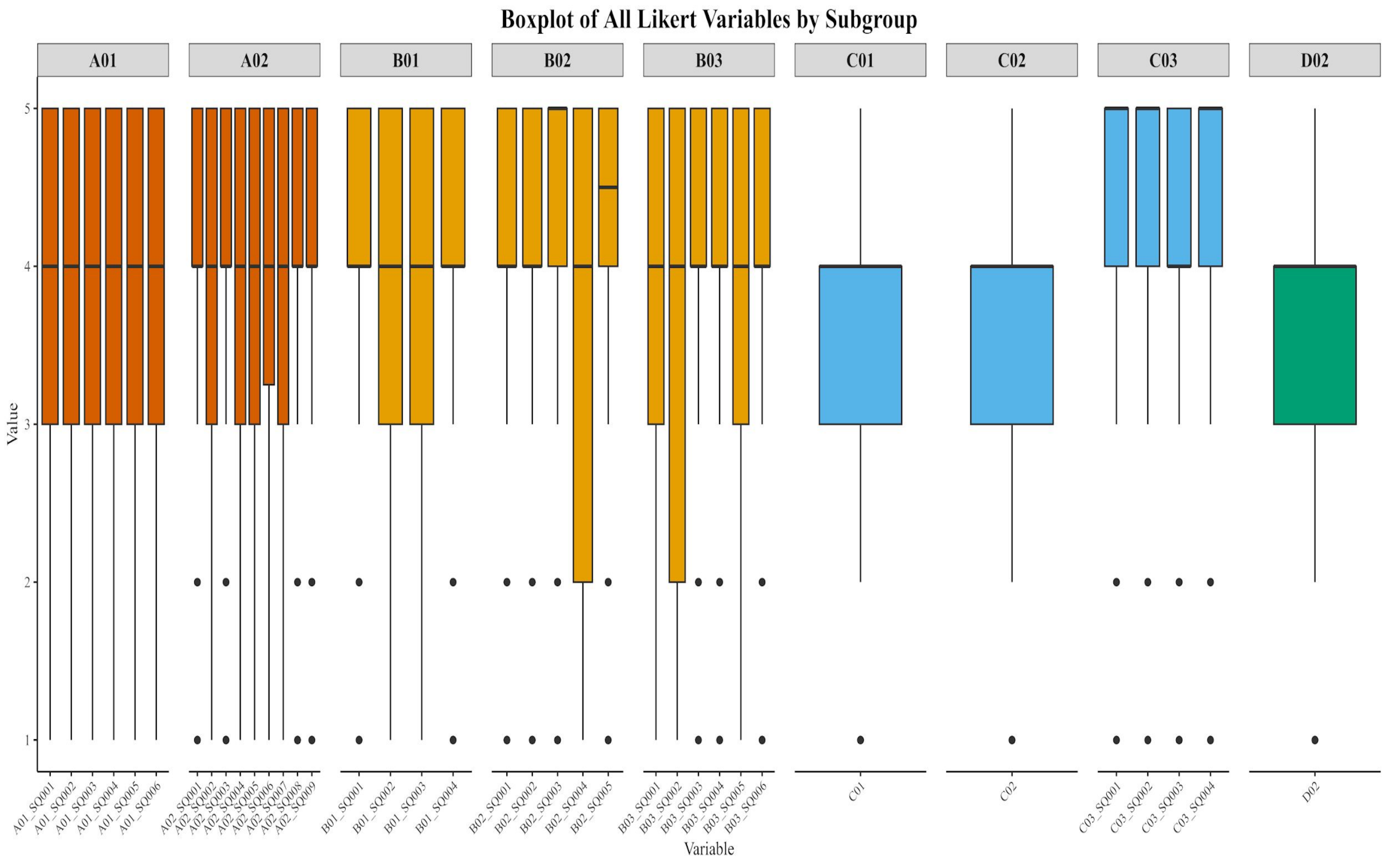
- The top five countries—**China, India, the United States, Poland**, and **South Korea**—together comprise a large portion of the total distribution
- Other countries like Taiwan, Brazil, Spain, Germany, and Italy also contribute notable numbers, but to a lesser extent
- A broad and diverse global spread

Distribution of Primary Focus Area



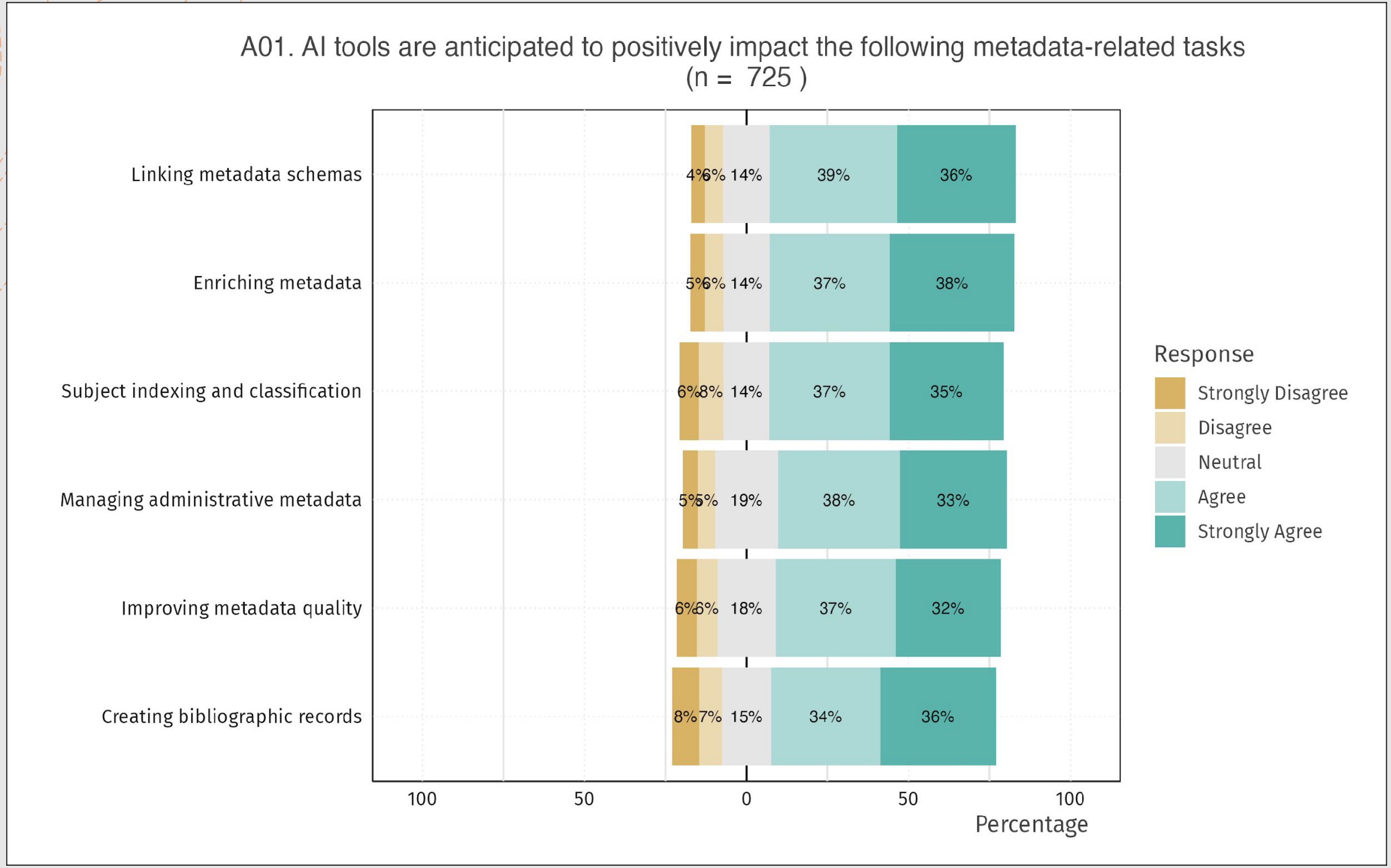
- The largest focus area is **Information Organization and Access**, which constitutes 40.4% of the total
- **User Services and Reference** accounts for 15.3%, and **Digital and Technology Services** comprises 13.9%
- Other: Engagement with AI, enthusiasm for its potential to streamline library operations and concerns about its impact on data integrity and employment within the sector

Distribution of Responses to Question Items



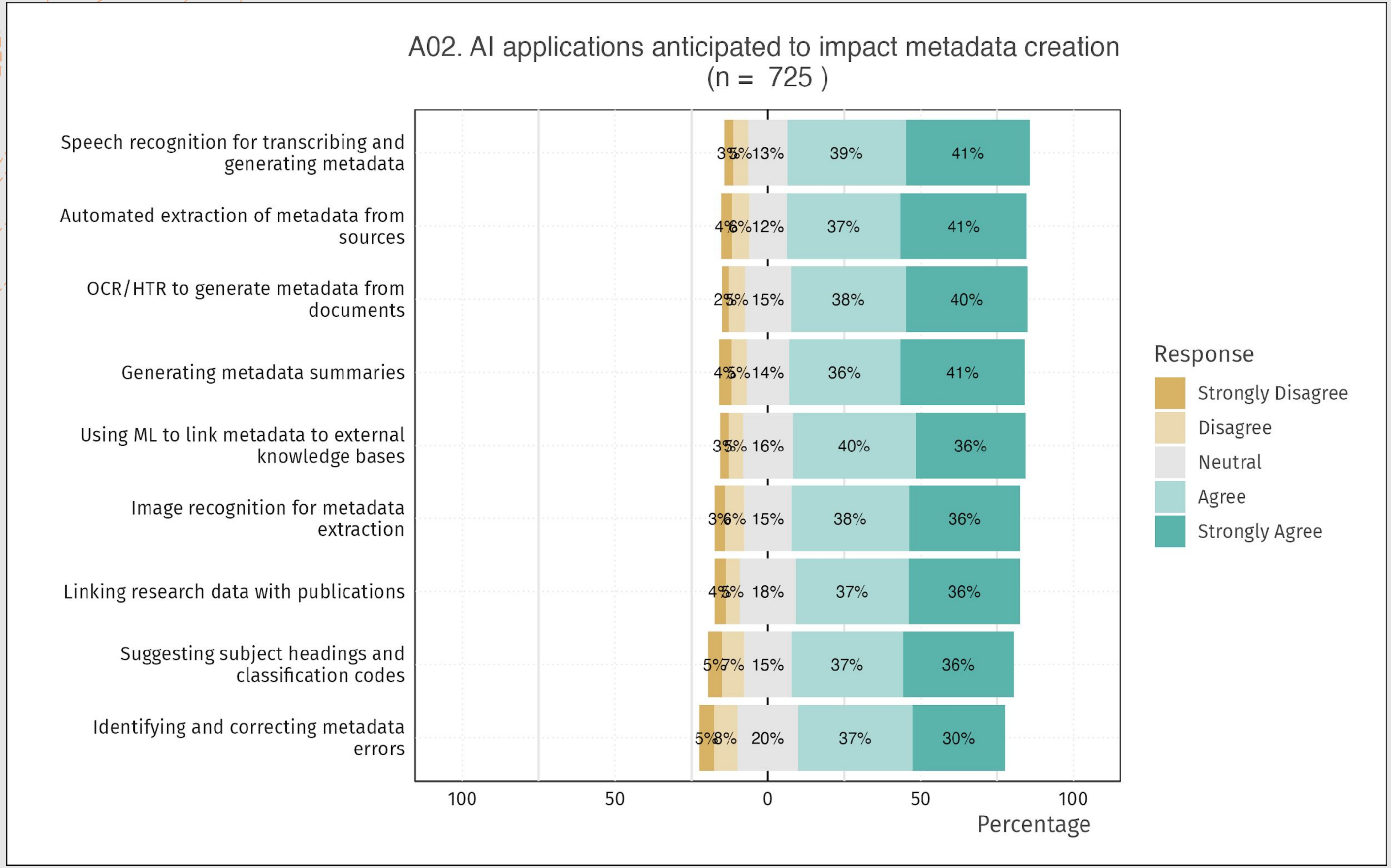
- **Group A** (A01: Impact on metadata tasks, A02: AI applications) generally positive responses
- **Group B**: B01: Benefits similar to Group A; B02: Challenges and B03: Concerns show greater variability
- **Group C**: High variability in C01 and C02 Beliefs about predictive AI and generative AI respectively; C03 Competencies generally positive
- **D02**: Confidence, a wide range of responses

A01: Anticipated Impact on Metadata Tasks



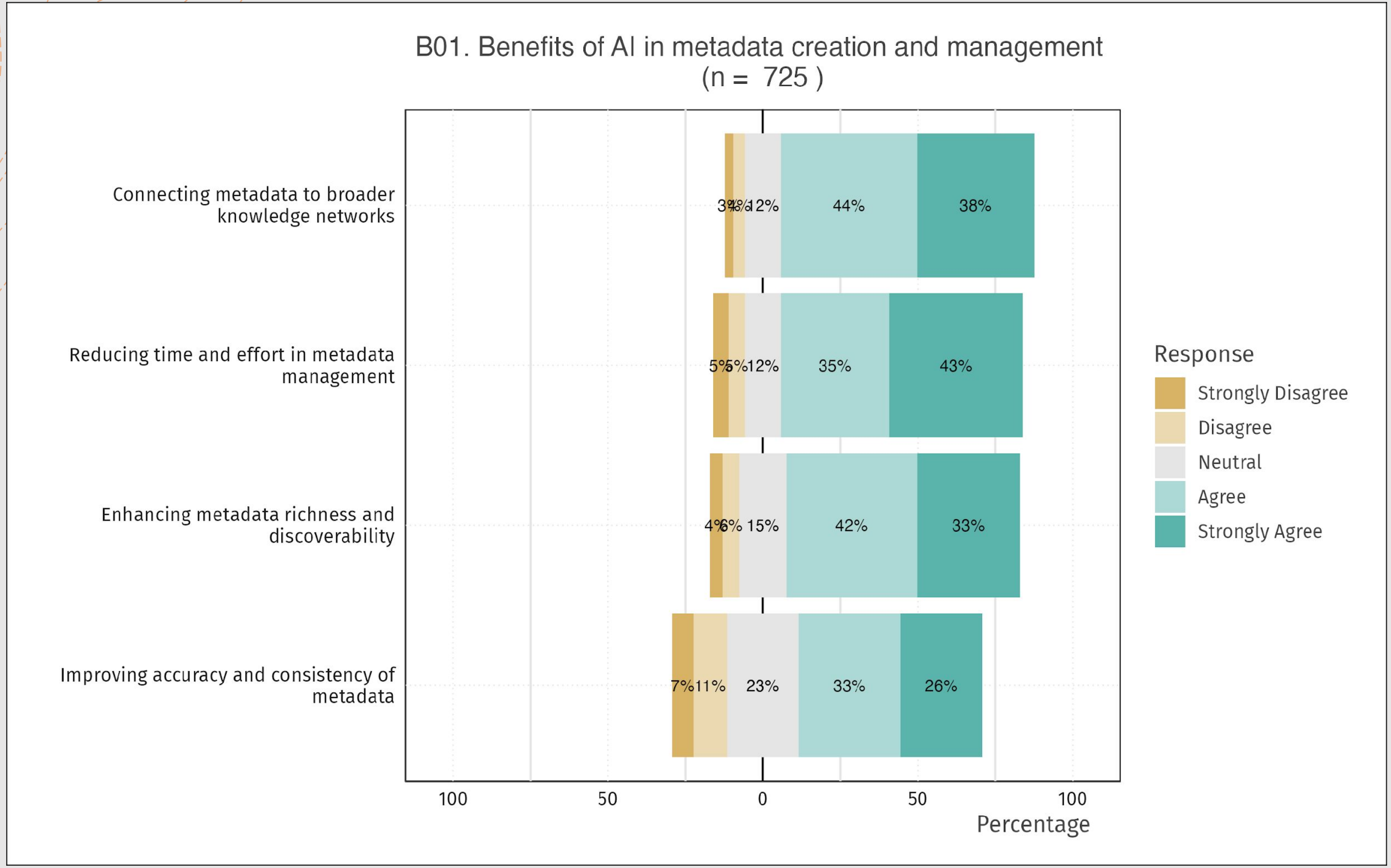
- Positive Impact: Most participants believe AI tools will positively impact metadata-related tasks
- High Agreement: Tasks like "Enriching metadata" and "Linking metadata schemas" have high agreement
- Overall Confidence: There is overall confidence in the benefits of AI for managing and improving metadata tasks

A02: Anticipated Impact on AI Applications



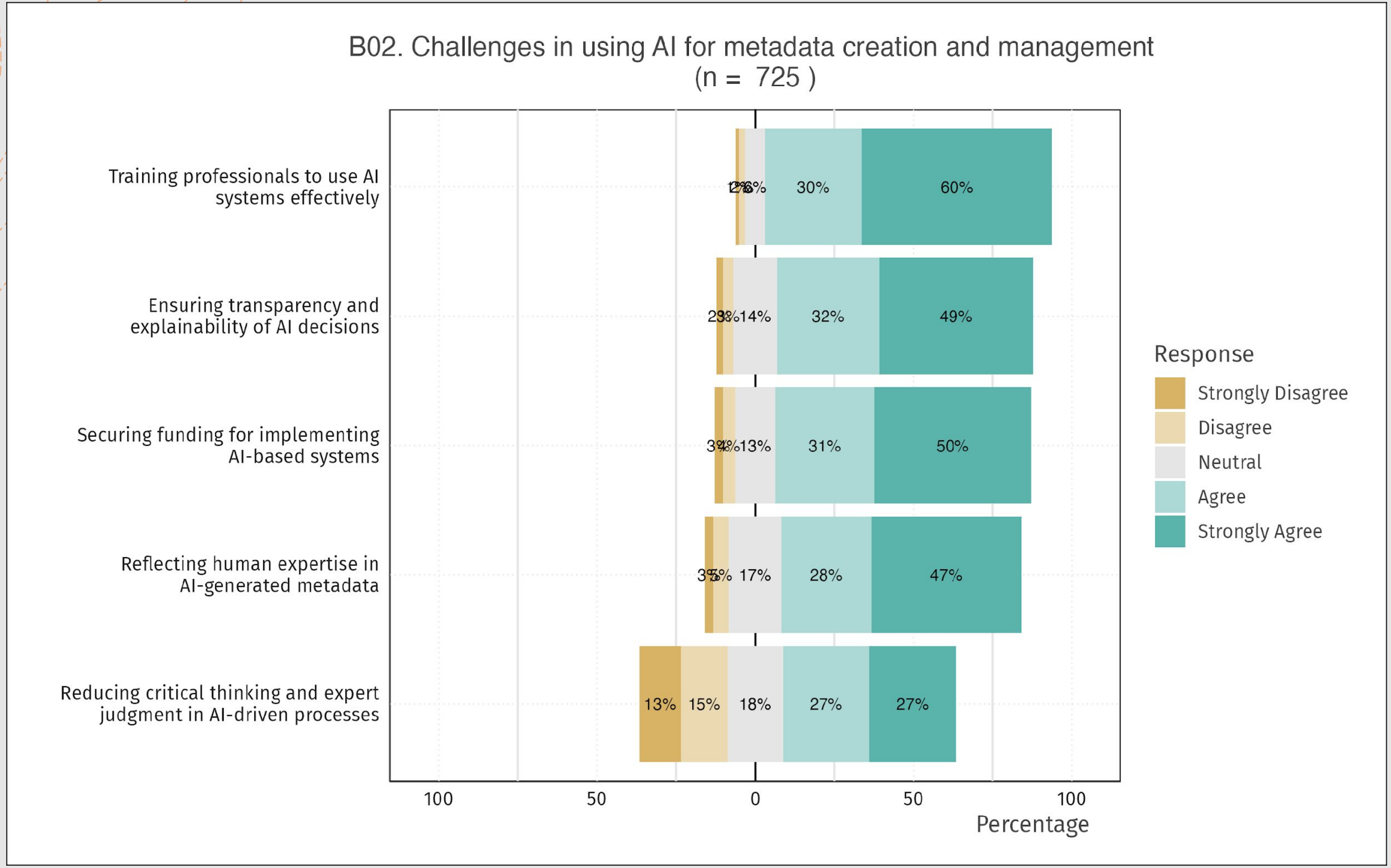
- Strong Confidence in AI: Participants have high confidence in AI's positive impact on metadata creation, such as speech recognition and metadata extraction
- Broad Support for AI Applications: There is strong support for various AI applications, including transcription, metadata extraction, and linking metadata to external knowledge bases

B01: Perceived Benefits



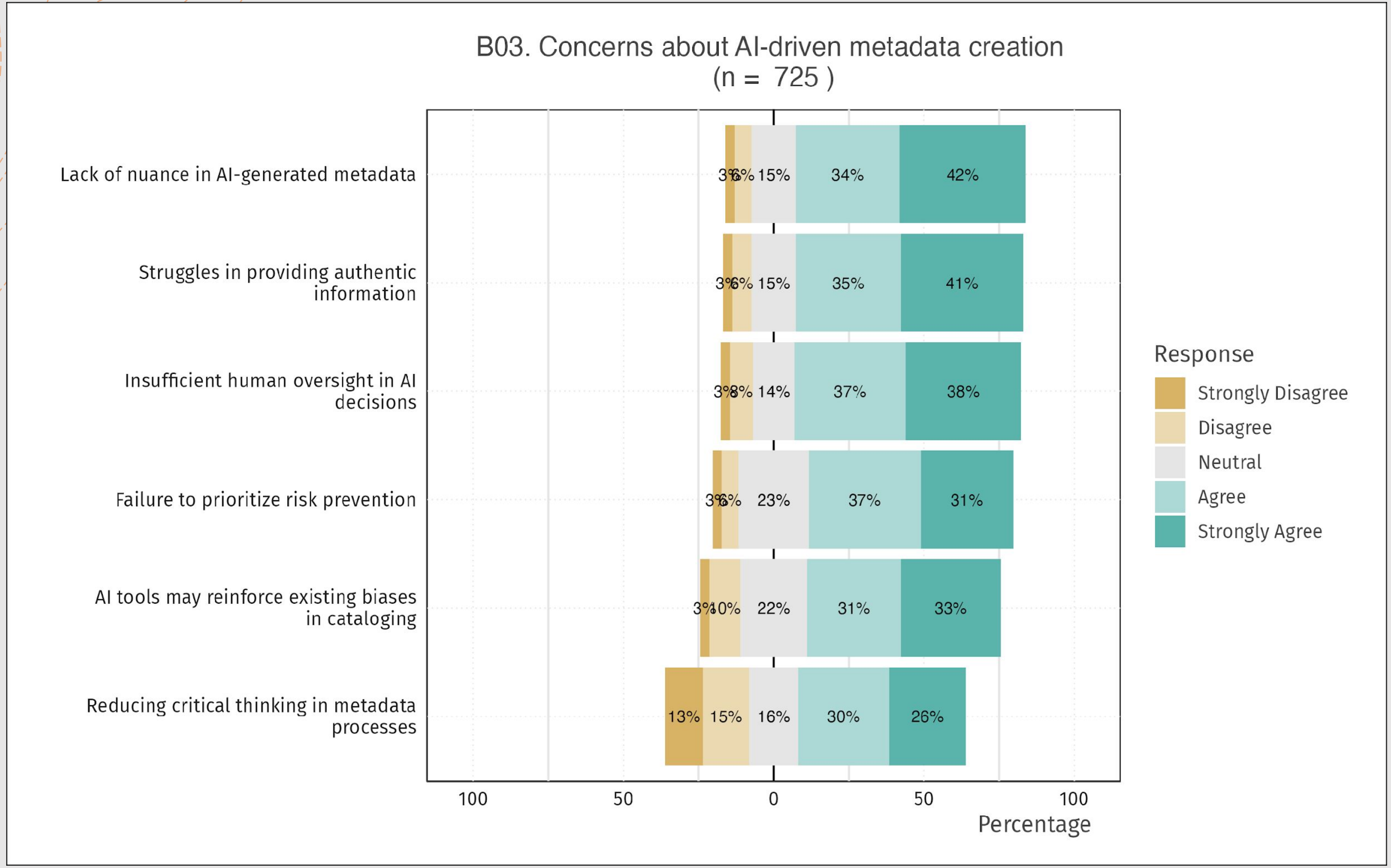
- High agreement on connecting metadata, reducing effort and enhancing metadata richness and discoverability
- Improving accuracy: While still positive, there's a slightly lower consensus on AI's impact on improving accuracy and consistency of metadata

B02: Challenges



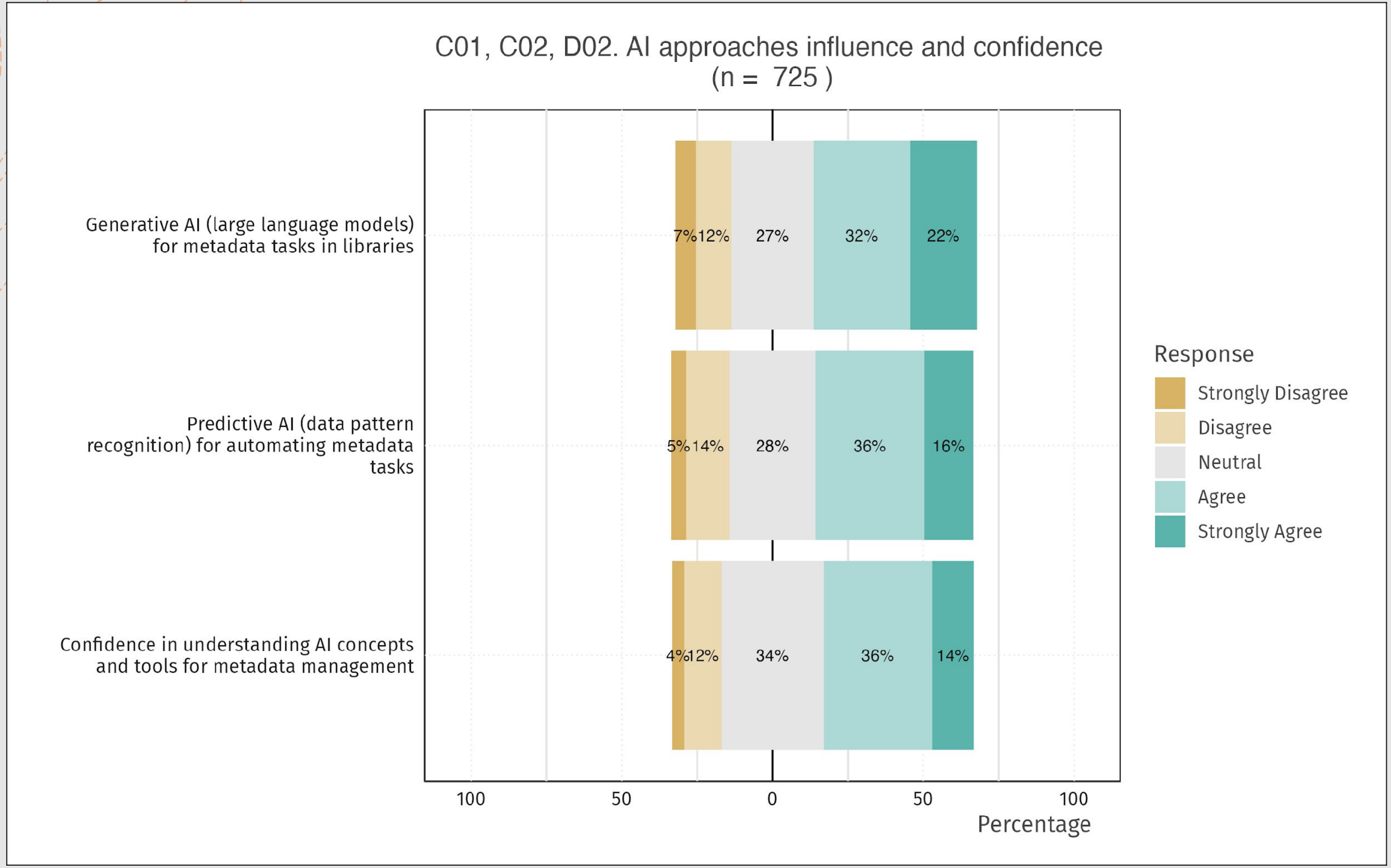
- Major challenges: training professionals, need for transparency, funding issues and integrating human expertise
- The responses reflect a cautious approach to integrating AI, emphasizing the need to strike a balance where AI enhances efficiency and accuracy while still valuing and incorporating human critical thinking and expertise

B03: Concerns



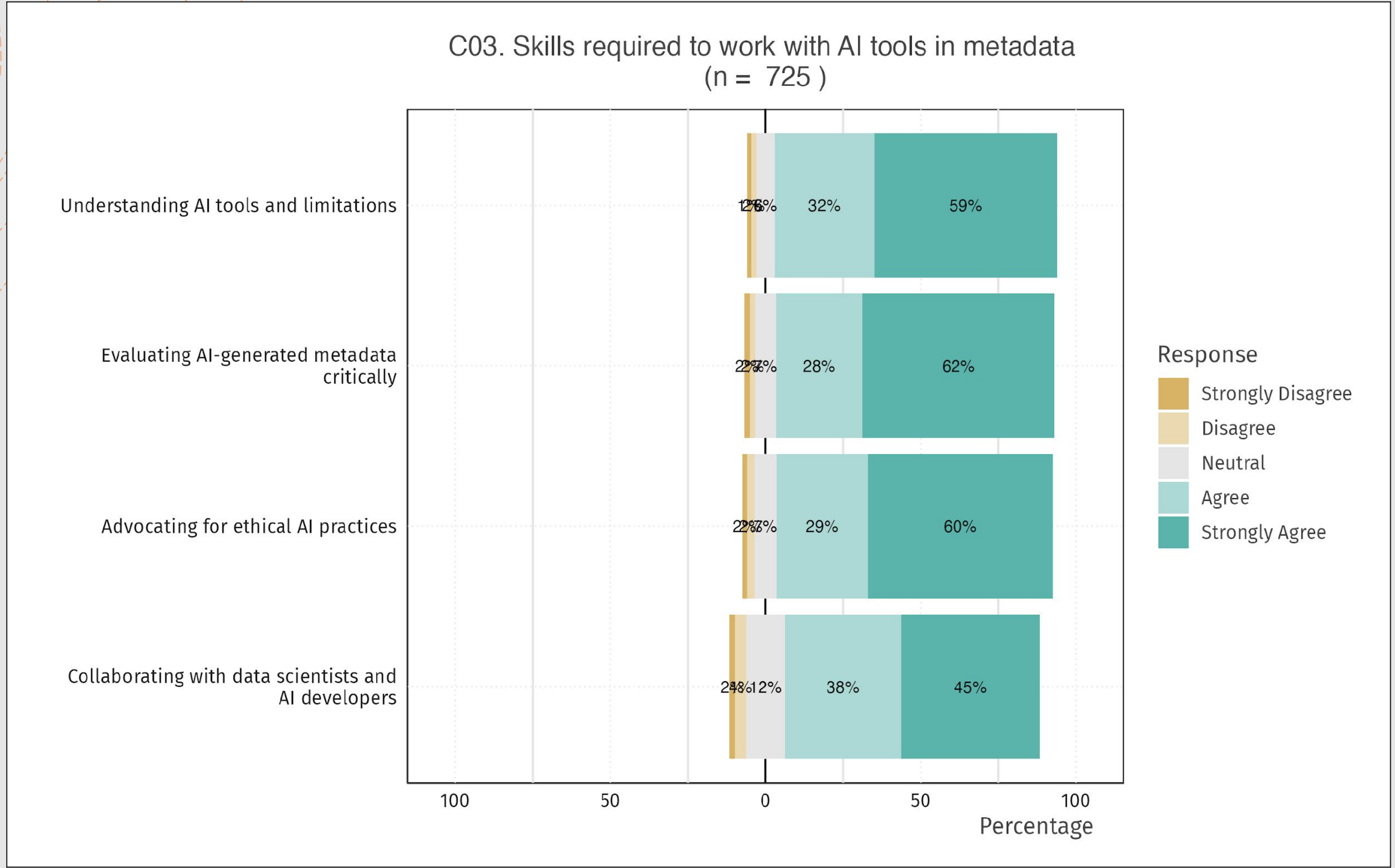
- Human Oversight: Crucial to maintain the quality and accuracy of AI-generated metadata
- Bias and Nuance: Concerns about AI tools reinforcing existing biases and lacking the subtlety and nuance of human-generated metadata
- Training and Transparency: Effective training for professionals and ensuring transparency and explainability of AI decisions

Anticipated AI Impact and Confidence



- A strong positive sentiment towards the use of both generative and predictive AI for metadata tasks; a notable minority (about 20%) with concerns or skepticism about their effectiveness
- Half of the participants believe they have a good grasp of how AI works and can effectively use AI tools in their metadata tasks

C03: Competency and skills



- A strong agreement on the key skills of understanding AI tools, critically evaluating AI outputs, advocating for ethical practices, and collaborating effectively with AI developers

CFA (Confirmatory Factor Analysis)

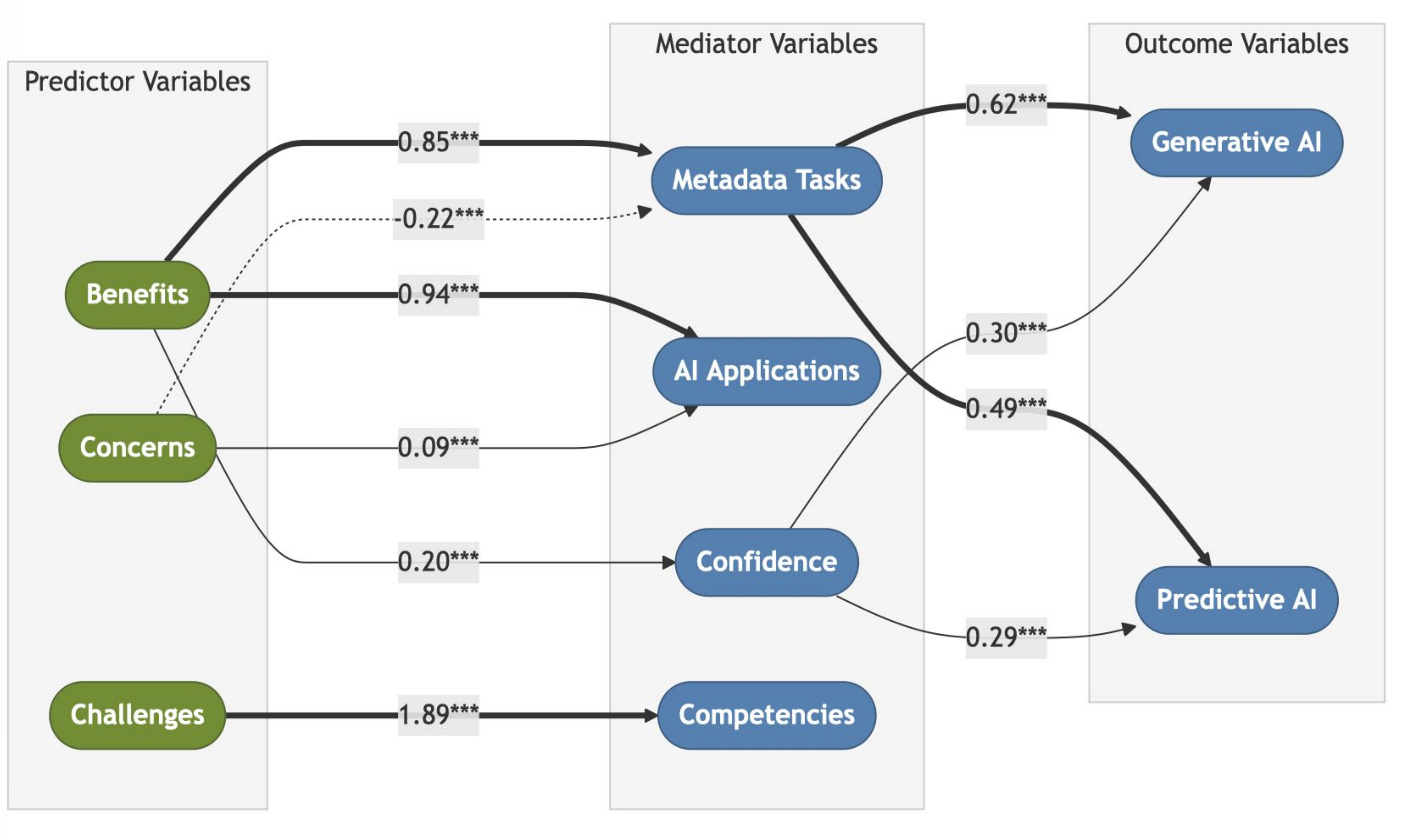
Model	χ^2	*df*	χ^2/df	*p*	CFI	TLI	RMSEA [90% CI]	SRMR
1-factor	6,092.58	77	79.12	< .001***	.72	.67	.33 [.32, .33]	.23
2-factor (B01 + B02, B03)	2,304.68	76	30.32	< .001***	.90	.88	.20 [.19, .21]	.16
2-factor (B01 + B03, B02)	5,645.60	76	74.28	< .001***	.74	.69	.32 [.31, .32]	.23
2-factor (B02 + B03, B01)	1,631.21	76	21.46	< .001***	.93	.91	.17 [.16, .17]	.13
3-factor	353.12	74	4.77	< .001***	.99	.98	.07 [.06, .08]	.06
Common guidelines[a]	**—**	**—**	**< 2 or 3**	**> .05**	**≥ .95**	**≥ .95**	**< .05 [.00, .08]**	**≤ .08**

[a]Based on Schreiber (2017), Table 3.

Fit indices for factor models: 1-factor, three 2-factor (varied combinations), and 3-factor models. Metrics: χ^2, df, χ^2/df, p, CFI, TLI, RMSEA (90% CI), and SRMR. Results suggest the 3-factor model demonstrates superior fit based on these indices, guided by Schreiber (2017).

- The 3-factor model provides the best fit for the data
- Benefits (B01), challenges (B02), and concerns (B03) related to AI and metadata represent distinct theoretical constructs
- Both challenges and concerns involve difficulties; challenges: opportunities for growth; concerns: potential problems and risks

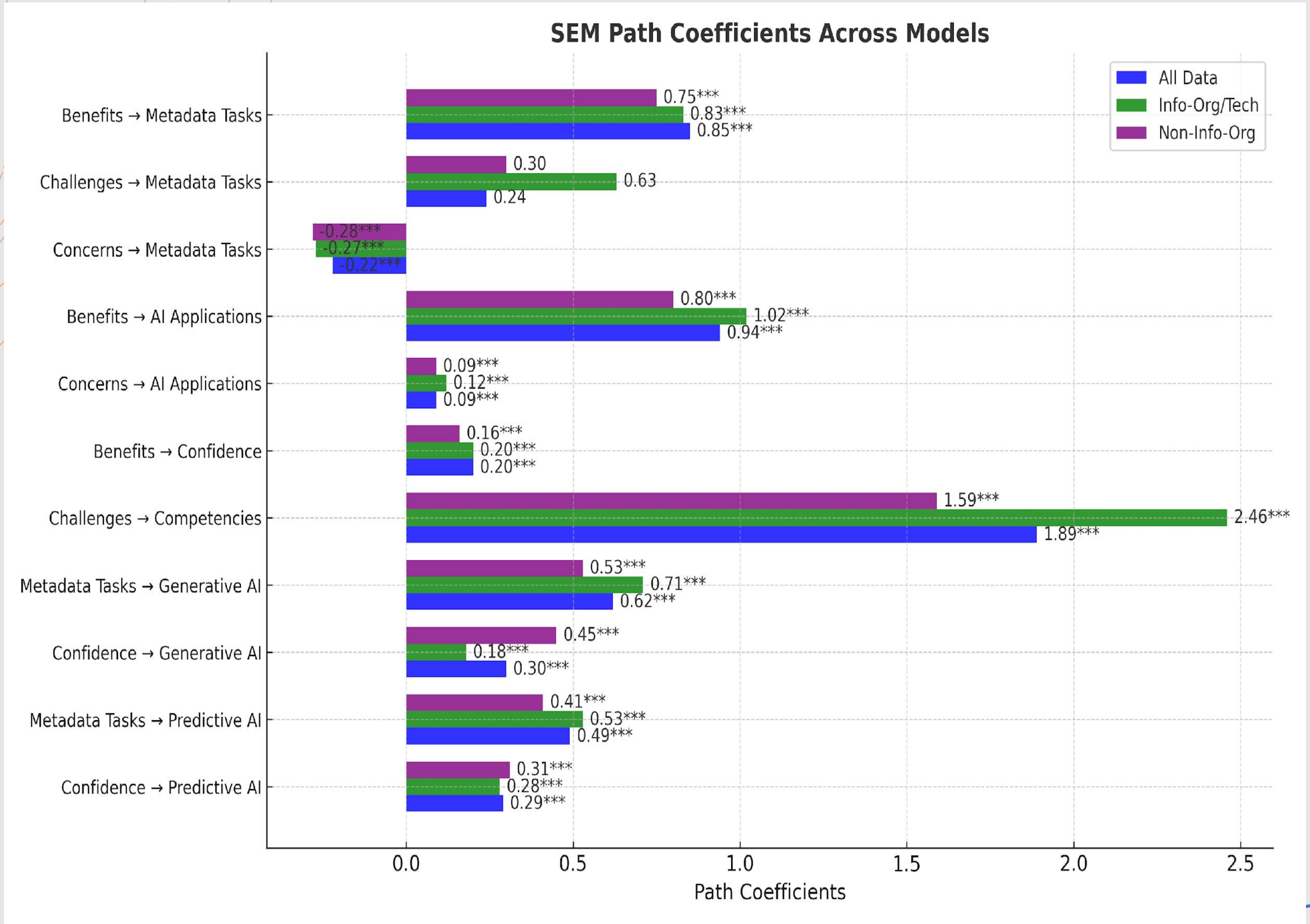
Structural Equation Modelling (SEM)



Structural Equation Modelling (SEM) of relationships among predictor variables (Benefits, Concerns, Challenges), mediator variables (Metadata Tasks, AI Applications, Confidence, Competencies), and outcome variables (Generative AI, Predictive AI). Path coefficients indicate the strength and significance of these relationships, highlighting key contributors to AI applications and competencies. *** indicates significant relationships.

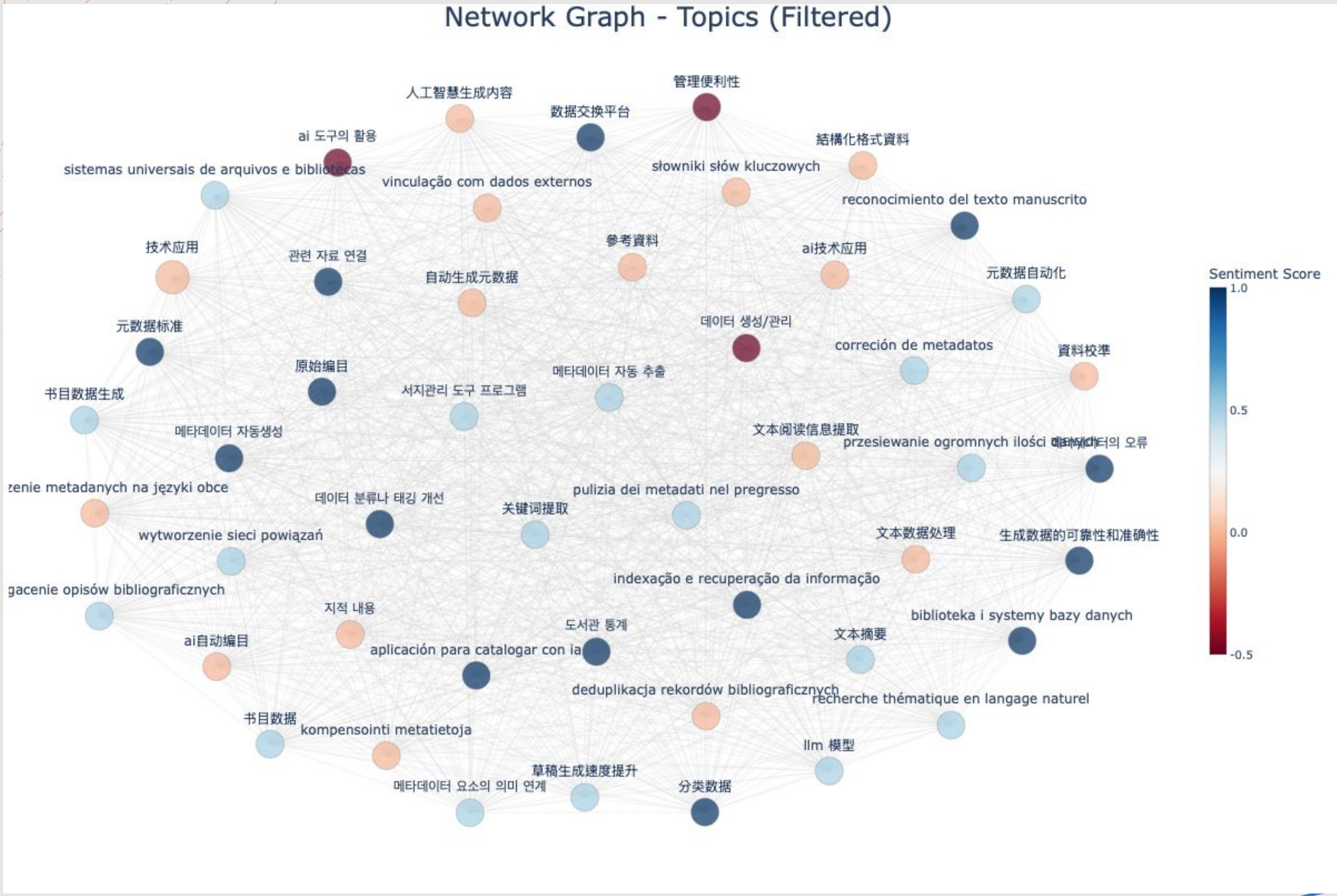
- Benefits strongly drive both Metadata Tasks and AI Applications
- Challenges significantly enhance competencies, suggesting the need for skills development
- Metadata Tasks is a key predictor of AI adoption for both Generative AI and Predictive AI
- Confidence is crucial for adopting AI, with a notable effect on both Generative and Predictive AI

Nested Models



- Challenges → Metadata Tasks: Significant in *Info-Org/Tech (0.63)**, but not in the other groups
- Challenges → Competencies: Highest effect in Info-Org/Tech (2.46*)**, showing that challenges strongly drive need for skill development
- Confidence → Generative AI: Strongest in Non-Info-Org (0.45*), weakest in Info-Org/Tech (0.18*)

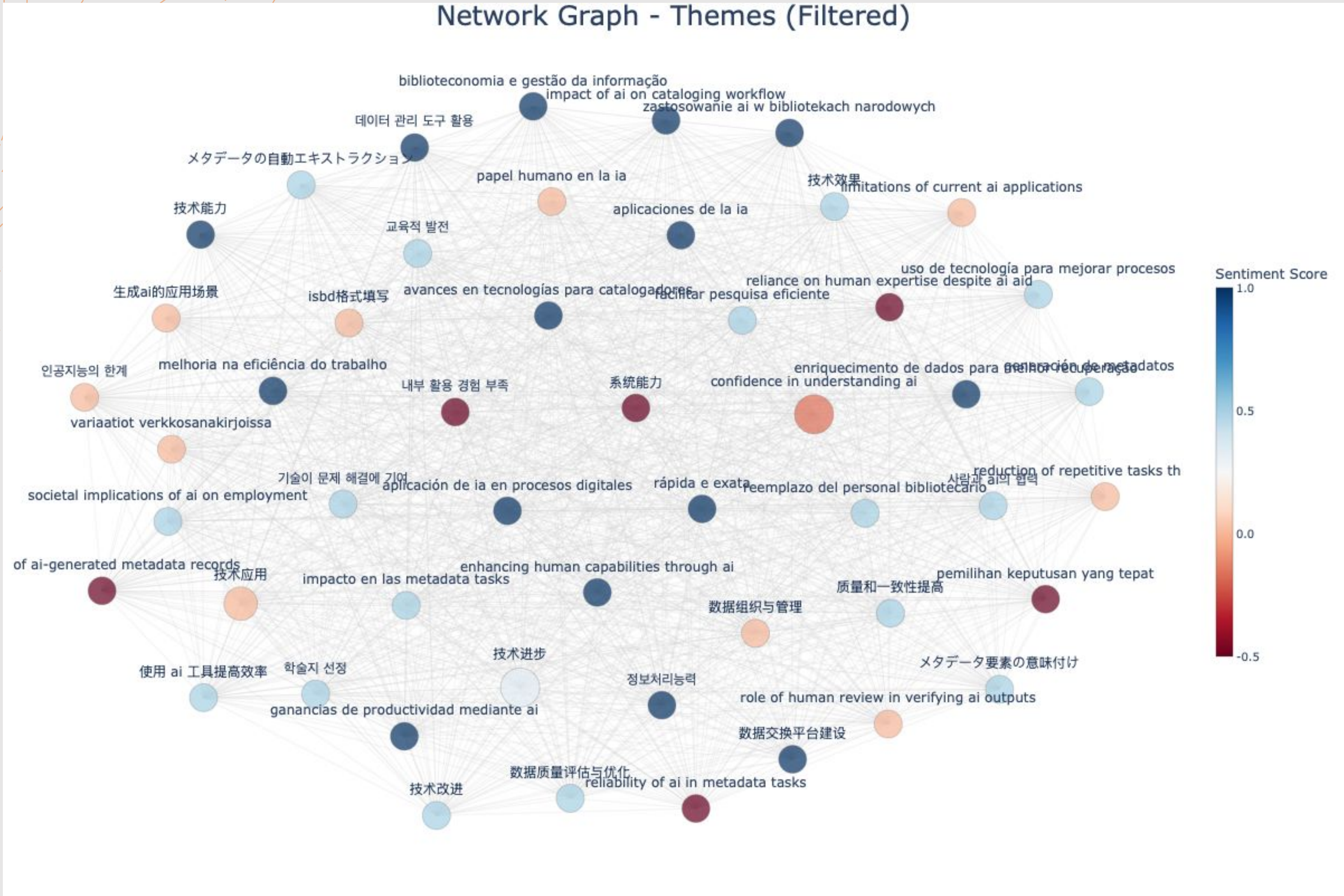
Network Graph of Well-Connected Topics



Top Topic Clusters

- Automatic/AI-based Metadata Generation
- Digitalization & OCR (Optical Character Recognition)
- AI-Based Solutions (NLP, Machine Learning, Classification/Indexing)
- Collaboration & Real-Time Metadata
- Integration with External Data Sources

Network Graph of Well-Connected Themes



1) Top Themes:

- Expanding human capabilities through AI
- Societal implications of AI on employment
- Impact in the metadata tasks
- AI for repetitive tasks in library processes
- AI-generated metadata creation

Key Competencies for Information Professionals

- **AI Fundamentals & Ethics**
  - Understanding AI tools, algorithms, and limitations.
  - Emphasis on AI ethics, bias detection, and climate concerns
  - Advocating for responsible AI practices
- **Technical & Data Skills**
  - Need for programming, scripting, and library management skills
  - Core data handling: data mining, indexing, quality assessment, cleaning
  - Data management & interoperability: **Standards** and system integration
- **Evaluation & Quality Assurance**

  Assessing AI-generated metadata for accuracy and bias

  Ensuring human verification for quality control

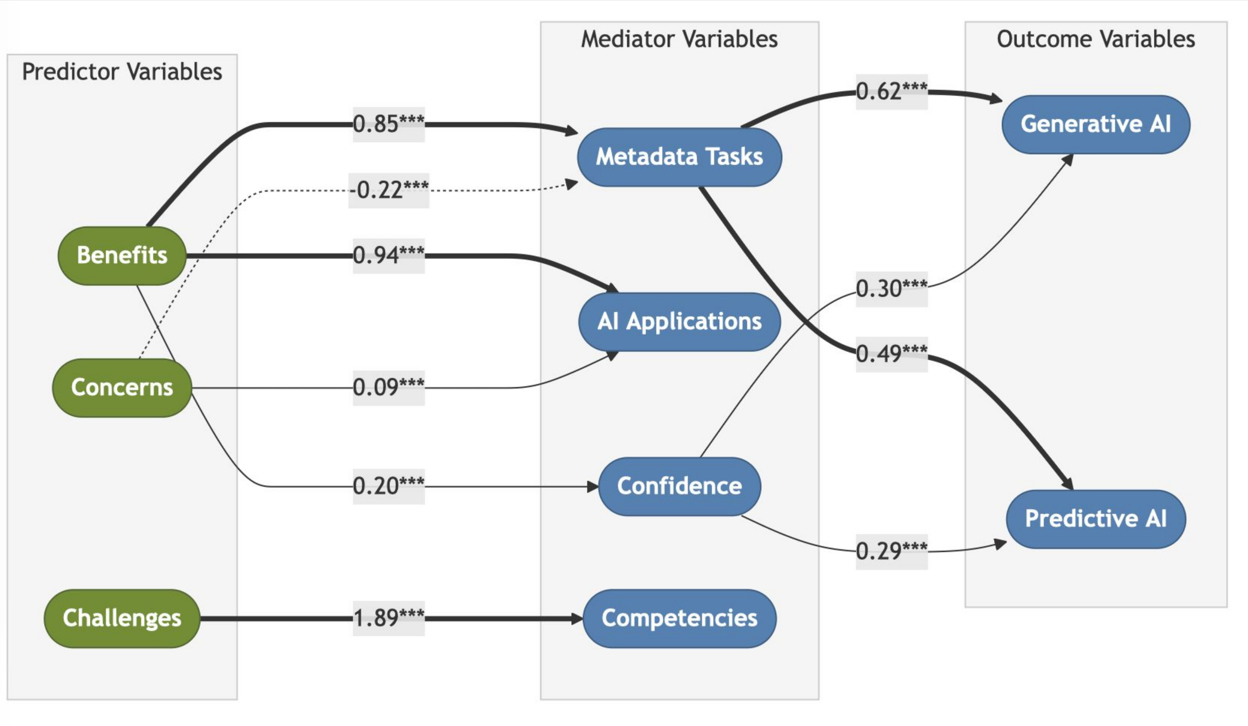
Metadata Task and AI-Related Tools

Metadata Task	AI and Related Tools
Metadata Creation & Generation	ChatGPT, AI MD-editor, OCR, AI for metadata from spreadsheets, images, voice, Small language models
Metadata Extraction	Grobid, OCR + NLTK, ABBYY FineReader, Transkribus, AI-powered NLP
Metadata Summarisation	ChatGPT, Library Robot
Metadata Classification & Tagging	Google Cloud Vision, Clarifai, AI for subject indexing & classification
Metadata Standardisation & Enrichment	DeepL, Google Translate, AI for schema reconciliation, spell-checking
Metadata Interoperability & Linking	Semantic retrieval discovery systems, Linked data environment, Annif tool
Metadata Quality Control	AI-driven quality checks, deduplication, disambiguation, Primo by Ex Libris, Tableau + AI plugins
Library Management Systems	OCLC's AI metadata tools, Alma primo, Automated scripts & workflows

Summary

- **Opportunities & Benefits**
  - Enhances metadata richness and discoverability
  - Reduces manual effort and expands human capabilities

- **Challenges & Concerns**
  - Robust professional training and adequate funding
  - Transparent AI decision-making and mitigation of potential biases
  - Integration with human expertise is critical for quality and accuracy



- **Key Skills for Information Professionals**
  - AI fundamentals and ethics
  - Technical competencies and evaluation techniques
- **Confidence in using AI** is essential to drive its successful adoption


thank
you