

Exploring the Taxonomy of Survey Papers on Large Language Models Using Classical Machine Learning

Maqsudur Rahman^{1*}, Md. Shahjahan²

¹Department of ICT, Comilla University, Cumilla, Bangladesh

²Directorate of Information and Communication Technology, Information and Communication Technology (ICT) Division, Dhaka, Bangladesh

Email: *mrrajon@cou.ac.bd, shahjahan.ict@gmail.com

How to cite this paper: Rahman, M. and Shahjahan, Md. (2025) Exploring the Taxonomy of Survey Papers on Large Language Models Using Classical Machine Learning. *Journal of Intelligent Learning Systems and Applications*, 17, 68-76.
<https://doi.org/10.4236/jilsa.2025.172006>

Received: December 1, 2024

Accepted: April 10, 2025

Published: April 30, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rapid advancements in large language models (LLMs) have led to an exponential increase in survey papers, making it challenging to systematically track and analyze their evolving taxonomy. This study employs graph representation learning combined with classical machine learning techniques to model and interpret the structural evolution of LLM-related survey papers. By constructing attributed graphs that capture topic distributions and interconnections, we provide a data-driven framework to explore research trends in this domain. A dataset of 241 survey papers published between July 2021 and January 2024 is analyzed to identify thematic developments and interdisciplinary relationships. The results highlight key areas of specialization, including the emergence of prompting science, multimodal models, and domain-specific applications in finance, education, and law. Co-occurrence analysis of survey topics reveals strong interconnections between core LLM research and fields such as software engineering, hardware architecture, and evaluation methodologies. These findings demonstrate the increasing specialization of LLM research and its growing integration across multiple disciplines. By leveraging graph-based methodologies, this study offers a structured approach to understanding the LLM survey landscape, facilitating efficient navigation of existing literature and identification of emerging research directions. The insights presented contribute to a more comprehensive understanding of the field's trajectory, assisting researchers and practitioners in engaging with the latest developments in LLM research.

Keywords

Large Language Models (LLMs), Taxonomy, Multimodal Models

1. Introduction

AI techniques have been widely applied to various domains, such as images [1] [2], texts [3] [4], and graphs [5] [6]. As a critical subset of AI techniques, Large Language Models (LLMs) have gained significant attention in recent years [7]-[12]. Especially, more and more new beginners are interested in the research topics about LLMs. To learn the recent progress in this field, new beginners commonly will read survey papers about LLMs. Therefore, to facilitate their learning, numerous survey papers on LLMs have been published in the last two years. However, a large amount of these survey papers can be overwhelming, making it challenging for new beginners to read them efficiently. To embrace this challenge, in this project, we aim to explore and analyze the metadata of LLMs survey papers, providing insights to enhance their accessibility and understanding [13].

Specifically, we aim to employ graph representation learning to analyze the taxonomy of survey papers on LLMs, focusing on both topical coverage and temporal trends. By leveraging this methodology, we aim to provide insights into the current state of the field and how it is evolving, helping researchers and practitioners navigate the complex landscape of LLM-related literature. Overall, our contributions can be summarized as follows:

- We introduce a graph-based framework to model and analyze the taxonomy of survey papers on LLMs, providing a scalable approach for understanding the structure of research topics.
- We offer a detailed temporal analysis of the growth in survey paper publications, highlighting key periods of increased activity and shifting research focus.
- Through our analysis, we identify significant trends in LLM-related research, including emerging areas such as “Prompting Science” and the application of LLMs in “Finance” and “Education.”
- By constructing and visualizing the interconnections between survey topics, we provide a clear understanding of the relationships between different research domains, enabling better navigation of the literature.

2. Methodology

In this section, we begin by outlining the data collection process, followed by an analysis of the metadata. Next, we describe how three types of attributed graphs are constructed and explain how graph representations are learned using graph neural networks.

2.1. Data Exploration

The dataset for this study comprises survey papers on Large Language Models (LLMs) published between July 2021 and January 2024. These papers were curated from publicly available research databases, primarily arXiv, along with select surveys from other open-access repositories. The inclusion criteria required that papers:

- Explicitly identify themselves as survey papers, review papers, or tutorials.

- Provide a structured discussion of previous research rather than presenting novel experimental results.
- Focus primarily on LLMs or their applications across various domains (e.g., prompting science, evaluation, adaptation techniques, and multimodal learning).

Exclusion criteria were applied to filter out:

- Opinion pieces, blog-style articles, and non-systematic reviews lacking structured literature analysis.
- Redundant or duplicate survey papers covering similar content without significant updates.
- Non-English surveys or those primarily discussing general AI trends without a distinct focus on LLMs.

After applying these criteria, we curated a dataset of 241 survey papers.

For each selected paper, we extracted key metadata attributes, including publication date, research topics, and author-provided keywords. This metadata was used to construct the graph representation for our analysis, where nodes represent research topics and edges denote relationships between topics based on co-occurrence within the same survey paper. This structure enables a systematic exploration of the evolving taxonomy of LLM-related research.

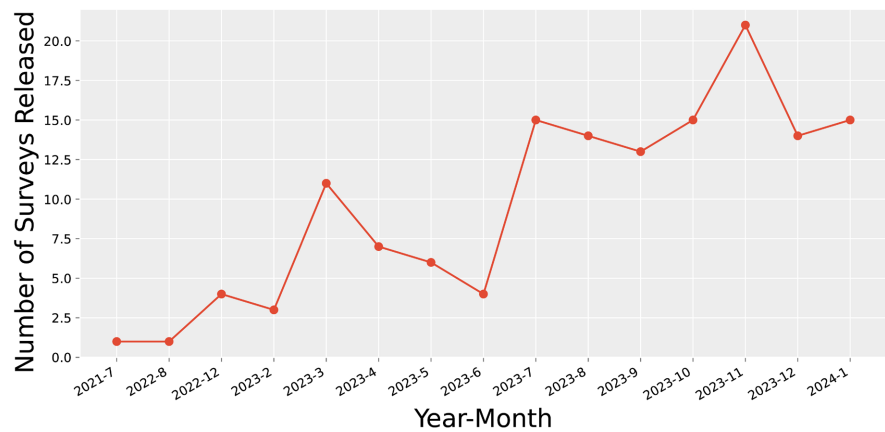


Figure 1. Trends of survey papers about large language models since 2021. Numbers reflect the year and month (e.g., 2023-3 is March 2023). Temporal distribution of survey papers released over the past years. The graph illustrates the increasing trend in the number of surveys published per month, highlighting the growing interest and research activity in the field.

2.2. Data Manipulation

In this phase, we focus on preparing the dataset by creating three graph types: text, co-author, and co-category. We build a TF-IDF matrix based on word frequency and distinctiveness for the titles and abstracts. To enrich the data, one-hot encoding is applied to arXiv categories, and this information is combined to generate the final feature matrix used for graph-based classification tasks.

The results from our graph-based analysis provide several key insights into the

taxonomy and trends of survey papers on LLMs: The number of surveys released has grown consistently from mid-2021 to late 2023, reflecting the rapidly expanding interest in LLM research. A sharp increase was observed from early 2022, peaking in mid-2023, as shown in the monthly publication trends (Figure from file [7]). This suggests an accelerated pace of development in the field, driven by both academic research and practical applications.

The categorization of survey papers shows a diverse set of research areas, with certain topics such as “Prompting Science” and “Evaluation” being particularly prominent (Figure from file [6]). This suggests that much of the current research is focused on understanding the best practices for prompting LLMs and evaluating their performance across various domains. Beyond core areas, we observed emerging interest in application-specific domains like “Finance,” “Law,” and “Education.”

These areas reflect the expanding use cases of LLMs in industry-specific applications, indicating a shift toward more practical, applied research.

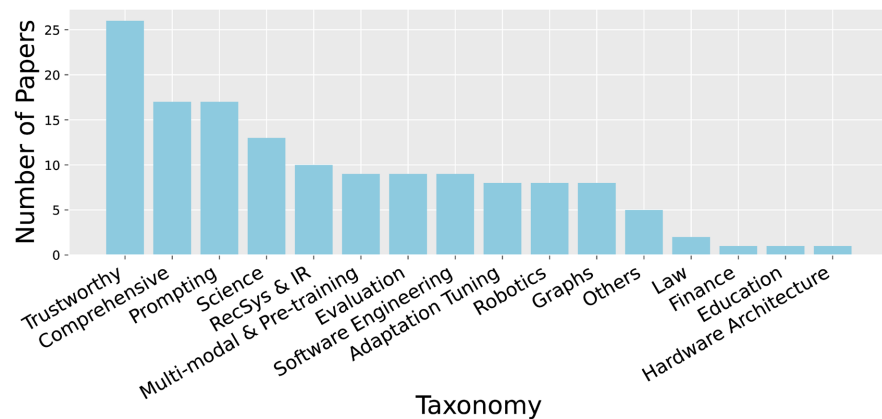


Figure 2. Categorization of research papers based on their thematic focus. The distribution showcases key research areas such as Trustworthy AI, Multi-modal Learning, Evaluation, and Robotics, indicating the diversification of topics within the surveyed domain.

The graph representation allowed us to explore the interconnections between topics. For example, surveys discussing “Multimodal Models” were often linked to “Pre-training” methodologies, highlighting the importance of foundational techniques in supporting new model architectures. Similarly, cross-disciplinary areas like “Software Engineering” and “Hardware Architecture” frequently intersected with core LLM research, underlining the multidisciplinary nature of the field.

2.3. Data Collection

The dataset consists of survey papers focused on large language models (LLMs) published between July 2021 and January 2024. The metadata for each paper includes the title, authors, publication date, keywords, and arXiv categories. Data sources: Survey papers were retrieved from publicly available databases such as

arXiv and other research repositories.

2.4. Graph Construction

To analyze the taxonomy and interconnections of survey papers on Large Language Models (LLMs), we constructed three types of attributed graphs, each capturing different aspects of research relationships.

Text Graph: This graph was built using a Term Frequency-Inverse Document Frequency (TF-IDF) matrix derived from the **title and abstract** of each survey paper. TF-IDF was used to quantify the significance of terms in individual papers relative to the entire dataset, allowing us to establish **semantic similarities** between papers. Nodes in this graph represent survey papers, while edges indicate high cosine similarity scores between their TF-IDF vectors, effectively grouping semantically related topics.

Co-Author Graph: In this graph, each node represents an **author**, and edges indicate **co-authorship relationships** between researchers. The edge weights are determined by the frequency of collaboration across multiple papers. This graph helps identify influential research communities, showing how expertise is distributed across different LLM-related domains.

Co-Category Graph: Nodes in this graph correspond to **arXiv categories**, and edges are established between categories that frequently appear together in the same survey paper. To **quantify** these interdisciplinary connections, we measured the co-occurrence of categories across the dataset, weighting edges based on the number of shared papers. This approach revealed key research overlaps, such as:

- Survey papers categorized under “*Software Engineering*” often discuss LLM deployment strategies, software optimization, and model efficiency, demonstrating strong links with “*Machine Learning*” and “*Artificial Intelligence*”.
- Papers under “*Hardware Architecture*” frequently reference discussions on computational scaling, transformer model parallelization, and quantization techniques, highlighting their intersections with “*Computational Complexity*”.

Feature Encoding: To further enrich the graph representations, we applied **one-hot encoding** to arXiv categories, converting categorical attributes into numerical features that were integrated into graph-based learning models. These encoded attributes improved the ability to cluster and classify research topics effectively.

By leveraging these three graph structures, we systematically modeled the relationships between survey papers, authors, and interdisciplinary research topics, enabling a comprehensive analysis of the evolving LLM research landscape.

2.5. Evaluation Metrics

To assess the effectiveness of the proposed graph-based methodology, we employed multiple evaluation metrics to ensure robust and reliable performance as-

assessment. The primary metrics used in this study include:

- **Classification Accuracy:** Measures the proportion of correctly classified instances among the total samples, providing a general assessment of model performance.
- **F1 Score:** A harmonic mean of precision and recall, particularly useful in scenarios where class imbalance is present.
- **Precision:** Evaluates the proportion of correctly predicted positive instances among all predicted positive instances, ensuring the reliability of classification outcomes.

To enhance generalizability and mitigate potential overfitting, we conducted a **5-fold cross-validation** procedure. The dataset was randomly partitioned into five subsets, with each fold used as a validation set while the remaining four served as the training set. This iterative process ensured that the model's performance was evaluated across multiple data splits, leading to a more **robust and unbiased** assessment.

The selected metrics and validation strategy provide a comprehensive evaluation framework, ensuring that the proposed approach effectively captures the taxonomy and evolving trends of LLM-related survey papers.

3. Results and Discussions

The number of surveys released has grown consistently from mid-2021 to late 2023, reflecting the rapidly expanding interest in LLM research. A sharp increase was observed from early 2022, peaking in mid-2023, as shown in the monthly publication trends (**Figure 1**). This suggests an accelerated pace of development in the field, driven by both academic research and practical applications.

Several factors contributed to this surge in LLM survey papers. First, the release of highly influential models, such as GPT-3, PaLM, and ChatGPT, sparked renewed research interest, leading to an increased demand for comprehensive literature reviews. Second, the growing accessibility of open-source LLMs, along with improvements in fine-tuning techniques, encouraged more domain-specific applications, necessitating surveys to consolidate best practices. Additionally, the emergence of new LLM capabilities, such as in-context learning, multimodal reasoning, and retrieval-augmented generation, led to distinct research directions, each requiring structured analysis. These technological and academic advancements collectively explain the observed rise in survey publications.

The categorization of survey papers shows a diverse set of research areas, with certain topics such as "Prompting Science" and "Evaluation" being particularly prominent (**Figure 2**). This suggests that much of the current research is focused on understanding the best practices for prompting LLMs and evaluating their performance across various domains. Beyond core areas, we observed emerging interest in application-specific domains like "Finance," "Law," and "Education." These areas reflect the expanding use cases of LLMs in industry-specific applications, indicating a shift toward more practical, applied research.

The graph representation allowed us to explore the interconnections between research topics. For example, surveys discussing “Multimodal Models” were often linked to “Pre-training” methodologies, highlighting the importance of foundational techniques in supporting new model architectures. Similarly, cross-disciplinary areas like “Software Engineering” and “Hardware Architecture” frequently intersected with core LLM research, underlining the multidisciplinary nature of the field. These connections were quantified through topic co-occurrence analysis in our constructed graph. Specifically, we observed that survey papers discussing efficient model deployment often shared categories with “Software Engineering” due to their focus on optimizing inference pipelines. Similarly, papers categorized under “Hardware Architecture” frequently referenced discussions on computational scaling, particularly regarding transformer model parallelization and quantization techniques.

4. Case Study: Evolution of “Prompting Science” as a Research Focus

To illustrate the impact of our graph-based methodology, we conducted a focused analysis on the development of “Prompting Science” as a research domain. In early survey papers (2021-2022), research was primarily centered on zero-shot and few-shot prompting techniques for text-based LLMs. However, from mid-2023 onwards, our taxonomy graph revealed a clear expansion into multimodal prompting, reflecting the rise of vision-language models and multi-agent collaborations. Additionally, survey papers in this category increasingly connected with “Evaluation” research, as benchmarking methodologies for prompting strategies became a critical subtopic.

These findings underscore the effectiveness of graph-based methodologies in capturing research evolution over time. By visualizing the growing interconnections between topics, we provide a structured approach to understanding how LLM-related research directions are emerging, merging, or diverging in response to technological advancements.

5. Conclusion

This study presents a graph-based approach to analyzing the taxonomy of survey papers on large language models (LLMs), offering a structured framework to understand the evolving research landscape. By leveraging graph representation learning and classical machine learning techniques, we systematically explored the interconnections between survey topics, their temporal evolution, and interdisciplinary relationships. Our findings highlight a clear trajectory of increasing specialization in LLM research. The results demonstrate significant growth in survey publications, with key trends emerging in areas such as prompting science, multimodal models, and application domains like finance, education, and law. Additionally, co-occurrence analysis revealed strong interdisciplinary links, particularly between LLM methodologies and fields such as software engineering, hard-

ware architecture, and evaluation frameworks. By modeling survey metadata as an attributed graph, this study enables researchers and practitioners to efficiently navigate the expanding body of literature. The proposed approach not only enhances the accessibility of LLM-related knowledge but also provides valuable insights into the structural evolution of the field. Future work can extend this analysis to incorporate evolving graph models and real-time updates, further refining our understanding of emerging research directions in LLMs.

6. Limitations and Future Scope

Despite the insights gained through this study, there are a few limitations to acknowledge: Our analysis is dependent on the availability and accessibility of survey papers. Some recent or domain-specific surveys might not be included due to dataset limitations. While graph representation learning effectively captures relationships between topics, the complexity of research topics can lead to oversimplification. Certain interdisciplinary areas may not be adequately captured by our clustering approach. The graph-based methodology is specifically tuned to analyze LLM-related survey papers. Its applicability to other fields of research remains to be tested and may require adjustments.

This project opens up several avenues for future work: Future studies can expand the dataset by including more diverse sources of survey papers, particularly those published in niche domains or non-English publications. Incorporating time-evolving graph models could help capture the dynamic evolution of research areas, enabling real-time updates of the taxonomy. Applying advanced natural language processing (NLP) techniques could allow for automated extraction of topics and relationships from survey papers, reducing manual curation efforts. The graph-based approach developed in this project could be adapted for analyzing survey papers in other fast-evolving fields such as artificial intelligence, computer vision, or bioinformatics.

Acknowledgements

The authors would like to express their gratitude to the research community for providing open-access survey papers that facilitated this study. We also acknowledge the contributions of anonymous reviewers for their constructive feedback, which helped improve the quality of this work. Additionally, we thank our colleagues for their insightful discussions and support throughout the research process.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>

- [2] Dosovitskiy, A. (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [4] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv: 1810.04805.
- [5] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. arXiv: 1609.02907.
- [6] Zhuang, J. and Al Hasan, M. (2022) Defending Graph Convolutional Networks against Dynamic Graph Perturbations via Bayesian Self-Supervision. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 4405-4413. <https://doi.org/10.1609/aaai.v36i4.20362>
- [7] Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I. (2018) Improving Language Understanding by Generative Pre-Training. OpenAI.
- [8] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019) Language Models Are Unsupervised Multitask Learners. OpenAI Blog, 1, 8, 9.
- [9] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G. and Askell, A. (2020) Language Models Are Few-Shot Learners. *Advances in Neural Information Processing Systems*, **33**, 1877-1901.
- [10] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S. and Anadkat, S. (2023) GPT-4 Technical Report. arXiv: 2303.08774.
- [11] Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A. and McKinnon, C. (2022) Constitutional AI: Harmlessness from AI Feedback. arXiv: 2212.08073.
- [12] Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M. and Hauth, A. (2023) Gemini: A Family of Highly Capable Multimodal Models. arXiv: 2312.11805.
- [13] Zhuang, J. and Kennington, C. (2024) Understanding Survey Paper Taxonomy about Large Language Models via Graph Representation Learning. *Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024)*, Bangkok, August 2024, 58-69. <https://doi.org/10.18653/v1/2024.sdp-1.6>