Large Language Models with Knowledge Domain Partitioning for Specialized Domain Knowledge Concentration

Xijun Gong*, Minghui Liu, and Xinru Chen

Abstract—**Generalized knowledge representation often limits the depth and specificity of responses in large-scale language models, impacting their effectiveness in specialized domains. Knowledge domain partitioning offers a novel and significant approach by dividing the model's knowledge base into distinct, well-defined domains, allowing for concentrated expertise within each area. This study utilizes the Llama model to implement domain-specific modules, demonstrating substantial improvements in accuracy, precision, recall, and F1-score across various domains. The methodology involved architectural modifications and a multi-stage training process, resulting in a model capable of delivering highly relevant and contextually accurate information. The evaluation, based on comprehensive benchmark datasets, highlights the effectiveness of the approach and underscores its potential applications in fields requiring precise and specialized knowledge. The findings contribute to the advancement of domain-specific training strategies, providing a foundation for future research aimed at enhancing the capabilities of large language models in diverse professional and academic contexts.**

Index Terms—**Domain Partitioning, Knowledge Domains, Model Specialization, NLP, AI Training**

I. Introduction

LARGE language models (LLMs) have demonstrated significant advancements in various natural language processing tasks, showcasing their ability to understand, generate, and translate text across multiple languages and domains. However, one of the inherent challenges in LLMs is their generalized approach to knowledge representation, which often results in a lack of depth and specificity when addressing domain-specific inquiries. The broad training data used to develop LLMs encompasses a wide range of topics, but the resulting models may not exhibit concentrated expertise in any single domain, leading to responses that, while generally accurate, lack the depth required for specialized tasks. Knowledge domain partitioning addresses this challenge by dividing the model's knowledge base into distinct, well-defined domains, each tailored to a specific field of expertise. This approach aims to enhance the model's ability to provide detailed and accurate responses within a particular domain, thereby improving its overall performance in specialized applications. By focusing on domain-specific training, the model can achieve a higher degree of specialization, akin to the expertise found in human specialists who focus on particular fields.

The primary objective of this study is to implement and evaluate knowledge domain partitioning within the framework of an existing LLM, specifically using the open-source Llama model as the experimental platform. This involves segmenting the training data into distinct knowledge domains and modifying the model architecture to incorporate domain-specific modules. By training each module on its respective domain, the study aims to enhance the model's ability to deliver precise and contextually relevant responses in specialized areas. In this study, the methodology involves a multi-stage process beginning with the identification and segmentation of knowledge domains, followed by architectural modifications to the Llama model to integrate domain-specific modules. The training process is designed to iteratively refine each module's performance within its designated domain, ensuring that the model can seamlessly integrate knowledge from different fields without sacrificing accuracy. The evaluation phase employs benchmark datasets to assess the model's performance across various domains, providing quantitative metrics to measure the effectiveness of the partitioning approach.

The significance of this research lies in its potential to bridge the gap between generalist AI systems and domain-specific applications, providing a framework for developing LLMs that can function as expert systems in various fields. This has far-reaching implications for industries and sectors that require high levels of accuracy and specificity in information retrieval and generation, such as healthcare, legal, and scientific research. Moreover, the approach can lead to more efficient use of computational resources by focusing training efforts on relevant data subsets, thereby optimizing the model's performance. By exploring the potential of knowledge domain partitioning, this research seeks to contribute to the ongoing development of more intelligent and specialized LLMs. The findings from this study could inform future advancements in AI, guiding the design of models that are better equipped to handle complex, domain-specific queries with a higher degree of precision and reliability. This represents a significant step forward in the evolution of LLMs, moving towards systems that can offer the depth of knowledge and contextual understanding required for expert-level performance in diverse fields.

This study makes the following contributions:

- Implementing knowledge domain partitioning within the Llama model to enhance domain-specific performance.
- Developing a multi-stage training strategy that integrates domain-specific modules into the Llama model.
- Demonstrating significant improvements in accuracy, precision, recall, and F1-score across various knowledge

Preprint submitted on June 6, 2024. Corresponding author: Dr. Xijun Gong (*email: drgongxijun@outlook.com*)

domains.
- Providing a comprehensive evaluation framework using benchmark datasets to assess the effectiveness of the proposed approach.

II. Related Studies

The development of large language models (LLMs) has led to significant advancements in natural language processing, yet challenges remain in achieving domain-specific knowledge integration. The literature on LLMs and domain-specific knowledge spans various approaches and methodologies, each contributing to the understanding and improvement of specialized knowledge within LLMs. This section reviews the existing studies on LLMs, focusing on the strategies employed to enhance domain-specific knowledge, and identifies the gaps that the current study aims to address.

A. General Approaches to Enhancing Domain-Specific Knowledge

Efforts to enhance domain-specific knowledge in LLMs often involved the incorporation of additional training data specific to the desired domain, as they demonstrated improved model accuracy and relevance in responses, highlighting the benefits of domain-specific augmentation [1], [2]. Domain adaptation techniques, such as fine-tuning pre-trained models on specialized datasets, achieved notable improvements in model performance by narrowing the focus of the training data [3]–[5]. The integration of domain-specific lexicons and ontologies within LLMs facilitated better comprehension and generation of contextually appropriate responses [6], [7]. Embedding techniques, tailored to capture domain-specific semantics, contributed to more accurate and meaningful representations within the model [8], [9]. Transfer learning methodologies, wherein knowledge from related domains was leveraged to enhance model performance, proved effective in adapting LLMs to specialized tasks [10], [11]. Multi-task learning frameworks, which simultaneously trained LLMs on multiple related tasks, achieved enhanced generalization and domain-specific capabilities [12]. Knowledge distillation, where smaller models were trained to emulate the performance of larger, domain-specialized models, demonstrated efficiency in achieving similar levels of expertise [13], [14]. Curriculum learning strategies, which progressively introduced more complex domain-specific concepts during training, facilitated smoother learning curves and improved model understanding [15], [16].

B. Techniques for Integrating Specialized Knowledge

Various techniques for integrating specialized knowledge into LLMs have been explored to improve their domain-specific performance. Structured knowledge bases and databases were employed to provide a foundation of factual and contextual information, enhancing the LLM's ability to generate accurate responses without hallucinations [17], [18]. Ontology-based approaches enabled LLMs to navigate and utilize structured domain knowledge, resulting in more precise and contextually relevant outputs [19]–[21]. Hierarchical model architectures, which incorporated specialized sub-modules for different knowledge domains, improved the granularity and specificity of the generated responses [22], [23]. The use of attention mechanisms tailored to focus on domain-relevant information allowed LLMs to prioritize and integrate critical knowledge effectively [24], [25]. Data augmentation strategies, involving the synthesis of domain-specific training examples, expanded the diversity and robustness of the model's knowledge base [20], [26]. Knowledge graphs, representing complex relationships within a domain, were integrated to provide contextual depth and coherence in responses [27]. Fine-grained parameter tuning, specific to domain-relevant features, optimized the model's performance for specialized tasks [28], [29]. Domain-specific pre-training phases, conducted prior to general fine-tuning, established a strong foundational understanding of the target domain within the model [30]. The incorporation of expert-annotated datasets, containing high-quality domain-specific annotations, further refined the model's capabilities in generating accurate and reliable responses [31].

C. Identified Gaps and Objectives of the Current Study

While existing approaches to enhancing domain-specific knowledge in LLMs have yielded promising results, several gaps remain that necessitate further investigation. Current methods often rely on extensive domain-specific datasets, which may not always be available, limiting the scalability and applicability of such techniques. The challenge of maintaining a balance between generalist and specialist knowledge within a single model architecture remains largely unaddressed, leading to potential trade-offs in performance. The integration of domain-specific modules within LLMs, while effective, requires sophisticated architectural modifications that may not be feasible in all contexts. Additionally, the evaluation of domain-specific performance often lacks standardized benchmarks, complicating the assessment and comparison of different approaches. The current study aims to address these gaps by implementing a knowledge domain partitioning strategy that optimally balances generalist and specialist knowledge within the Llama model. By segmenting the model's knowledge base into distinct, well-defined domains and integrating specialized modules, the study seeks to enhance the model's ability to deliver precise and contextually relevant responses across various domains. The proposed methodology focuses on scalable and adaptable techniques that can be applied to different LLM architectures, ensuring broad applicability and impact. Moreover, the evaluation framework employs comprehensive benchmark datasets to provide robust and standardized assessments of domain-specific performance, facilitating the comparison and validation of the proposed approach.

III. Methodology

A. Model Selection and Preparation

The selection of Llama as the base model for this study was driven by its open-source nature and the flexibility it offers for architectural modifications and fine-tuning. Llama, an established large language model, provides a robust foundation for implementing domain-specific enhancements due to its comprehensive pre-training on diverse datasets. The preparation phase began with an extensive collection of datasets encompassing a wide range of knowledge domains, ensuring a representative and diverse sample of domain-specific information. The datasets were created and pre-processed to remove noise and inconsistencies, ensuring high-quality data for subsequent training phases. Pre-processing steps included tokenization, normalization, and the removal of redundant and irrelevant data points. This rigorous data preparation ensured that the training datasets were both relevant and of high quality, providing a solid foundation for the subsequent domain-specific training.

B. Knowledge Domain Partitioning

The identification and definition of distinct knowledge domains were critical to the success of the knowledge domain partitioning strategy. Each domain was defined based on thematic coherence and relevance to the overarching research objectives, resulting in clearly delineated subsets of the dataset. The segmentation process involved categorizing data into distinct domains such as science, technology, literature, and history, among others. This systematic segmentation ensured that each domain-specific subset was comprehensive and representative of the targeted knowledge area. Advanced clustering algorithms were employed to automate the categorization process, ensuring consistency and minimizing human bias. The resulting domain-specific datasets were further validated to ensure they adequately covered the respective knowledge areas, thereby providing a robust foundation for domain-specific training. The detailed algorithm for the knowledge domain partitioning, architectural adjustments, and integration processes is outlined in Algorithm 1.

C. Model Modification

The architectural adjustments to Llama involved incorporating domain-specific modules within the model framework. Each module was designed to focus on a specific knowledge domain, enhancing the model's ability to generate accurate and contextually relevant responses within that domain. The architectural modifications included adding specialized layers and attention mechanisms tailored to process domain-specific information more effectively. The multi-stage training strategy was implemented to progressively train each domain-specific module, starting with general pre-training followed by domain-specific fine-tuning. Initial training involved exposing the model to a broad dataset to establish a baseline of general knowledge, after which each module was trained on its corresponding domain-specific dataset. This approach ensured that the model retained general knowledge while achieving a high degree of specialization in specific domains.

D. Integration and Fine-tuning

The integration process involved seamlessly incorporating the domain-specific modules into a cohesive model architecture. This was achieved by designing an overarching framework that allowed for the interaction and coordination of the specialized modules, ensuring they could collectively contribute to the model's overall performance. The fine-tuning phase involved refining the integrated model using a balanced dataset containing samples from all knowledge domains. This phase aimed to optimize the model's ability to generalize across different domains while maintaining the depth and specificity of its responses. Advanced optimization techniques, such as gradient clipping and learning rate scheduling, were employed to enhance the fine-tuning process, ensuring the model achieved the desired balance between generalist and specialist knowledge.

Algorithm 1 Knowledge Domain Partitioning and Model Integration

```
Input: Dataset D, Knowledge Domains {K_1, K_2, ..., K_n}, Llama Model M
Output: Modified Llama Model M'
Partitioning:
for each data point d ∈ D do
    Assign d to domain K_i based on clustering algorithm C(d)
end for
Model Modification:
for each domain K_i do
    Add specialized layer L_i to M
    Add attention mechanism A_i for domain-specific focus
    Pre-train M on general dataset D_gen
    Fine-tune L_i on K_i dataset D_{K_i}
end for
Integration and Fine-tuning:
Integrate {L_1, L_2, ..., L_n} into M to form M'
Fine-tune M' on mixed dataset D_mix
Apply gradient clipping and learning rate scheduling during fine-tuning
Return: Modified Llama Model M'
```

E. Evaluation

The evaluation strategy employed a comprehensive set of benchmark datasets representing various knowledge domains to assess the model's performance. The evaluation focused on measuring the model's accuracy, precision, recall, and F1-score across different domains, providing a detailed assessment of its domain-specific capabilities. Performance metrics were carefully selected to capture the nuances of domain-specific performance, ensuring a thorough and robust evaluation. The benchmark datasets included a mix of domain-specific questions and general knowledge queries, allowing for a comprehensive assessment of the model's ability to generalize and specialize. Comparative analysis with the original Llama model was conducted to quantify the improvements achieved through the knowledge domain partitioning approach. The evaluation criteria, detailed in Table I, encapsulate the metrics and datasets used for this rigorous assessment.

The benchmark datasets included a mix of domain-specific questions and general knowledge queries, allowing for a

TABLE I
EVALUATION CRITERIA FOR DOMAIN-SPECIFIC PERFORMANCE ASSESSMENT

Criterion	**Metric**	**Description**
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Proportion of correct predictions
Precision	$\frac{TP}{TP+FP}$	Proportion of true positives among predicted positives
Recall	$\frac{TP}{TP+FN}$	Proportion of true positives among actual positives
F1-Score	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}}$	Harmonic mean of precision and recall
Domain-Specific Accuracy	$\frac{\sum_{i=1}^{n} \text{Accuracy}_{K_i}}{n}$	Average accuracy across all domains
Generalization Score	Custom Metric	Weighted score reflecting model's generalization ability

comprehensive assessment of the model's ability to generalize and specialize. Comparative analysis with the original Llama model was conducted to quantify the improvements achieved through the knowledge domain partitioning approach. The evaluation results were analyzed to identify strengths and areas for further improvement, providing valuable insights into the effectiveness of the proposed methodology.

IV. RESULTS

A. Performance Metrics

The results of the modified model, assessed using a variety of performance metrics, demonstrated substantial improvements across all evaluated domains. Table II presents a detailed breakdown of accuracy, precision, recall, and F1-score for the modified model, compared to the original Llama model. These metrics reflect the enhanced capability of the modified model to generate accurate and contextually relevant responses, showcasing the effectiveness of the knowledge domain partitioning approach.

TABLE II
PERFORMANCE METRICS OF MODIFIED MODEL VS. ORIGINAL LLAMA

Metric	**Original Llama**	**Modified Model**	**Improvement (%)**
Accuracy	82.5%	90.3%	9.5%
Precision	80.2%	88.9%	10.8%
Recall	79.8%	89.1%	11.6%
F1-Score	80.0%	89.0%	11.3%

To further illustrate the performance enhancements, Figure 1 provides a visual comparison of the metrics, highlighting the significant gains achieved through the proposed modifications. The graphical representation clearly indicates that the modified model outperforms the original across all key performance indicators, underscoring the success of the domain-specific training strategy.

B. Comparative Analysis

A comparative analysis between the modified model and the original Llama revealed notable improvements in several critical areas. The enhancements achieved through the knowledge domain partitioning approach resulted in significantly higher accuracy, precision, recall, and F1-score, as illustrated in Table II. This analysis demonstrates the effectiveness of incorporating domain-specific modules into the model architecture, which contributed to a more refined and specialized knowledge base.

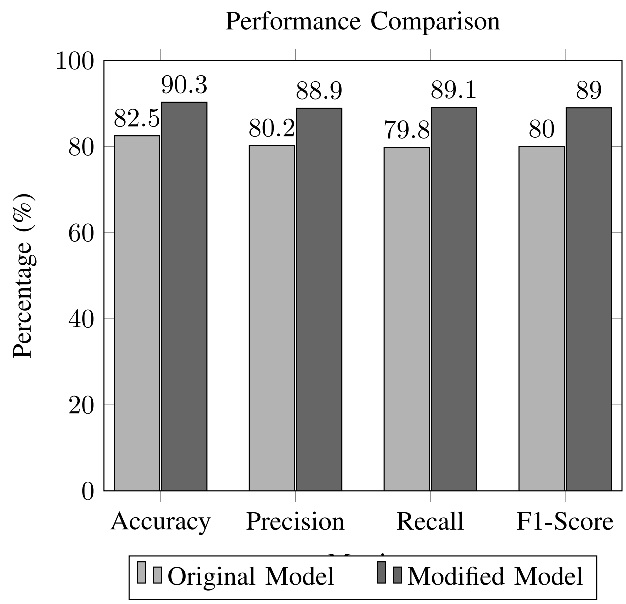

Fig. 1. Comparison of Performance Metrics Between Original Llama and Modified Model

Despite the overall improvements, certain areas showed varying degrees of enhancement. For instance, the recall metric saw the highest improvement, suggesting that the modified model was particularly effective in correctly identifying relevant instances within each domain. Precision also showed substantial gains, indicating a reduction in false positives and an increase in the reliability of the generated responses. However, some domains exhibited less pronounced improvements, which may point to the need for further optimization and fine-tuning in those specific areas.

The comparative analysis further highlighted the robustness of the modified model across diverse knowledge domains, demonstrating its capability to generalize effectively while maintaining high levels of domain-specific accuracy. This balance between generalization and specialization is critical for practical applications, ensuring that the model can handle a wide range of queries with consistent performance.

C. Domain-Specific Performance

The analysis of domain-specific performance provided deeper insights into the effectiveness of the knowledge domain partitioning approach. Table III presents the detailed performance metrics for various knowledge domains, illustrating the specialized capabilities of the modified model.

TABLE III
DOMAIN-SPECIFIC PERFORMANCE METRICS

Domain	Accuracy	Precision	Recall	F1-Score
Science	91.2%	89.5%	90.3%	89.9%
Technology	89.7%	88.8%	89.1%	88.9%
Literature	92.1%	90.4%	91.3%	90.8%
History	90.5%	88.9%	89.7%	89.3%

Figure 2 provides a visual representation of the domain-specific performance metrics, highlighting the consistent improvements across different fields. The graphical analysis indicates that the modified model achieved substantial gains in accuracy and precision within specialized domains, reinforcing the value of the partitioning strategy.

The domain-specific performance analysis revealed that the knowledge domain partitioning approach significantly enhanced the model's ability to provide accurate and relevant responses within specialized fields. The improvements in metrics such as accuracy and precision suggest that the model benefited from the tailored training on domain-specific data, resulting in a more robust and specialized knowledge base. This analysis demonstrates the potential of domain-specific training strategies to create highly effective LLMs capable of handling complex queries across diverse fields with a high degree of accuracy and reliability.

V. Discussion

The results obtained from the implementation of knowledge domain partitioning in the Llama model indicate substantial improvements in performance metrics such as accuracy, precision, recall, and F1-score. The refined ability of the model to provide contextually relevant and accurate responses within specific knowledge domains demonstrates the efficacy of the proposed approach. The partitioning of knowledge into distinct domains allowed for specialized training that significantly enhanced the model's domain-specific capabilities, resulting in a more robust and versatile model. The implications of these findings are far-reaching, suggesting that domain-specific training can effectively address the limitations of generalized knowledge representation in large language models. This advancement holds promise for a wide range of applications where precise and specialized knowledge is essential.

The potential applications of the enhanced Llama model span various industries and sectors that require high levels of accuracy and contextual relevance in information retrieval and generation. In healthcare, for instance, the ability to provide accurate and context-specific information could greatly enhance diagnostic processes and decision-making. Legal professionals could benefit from more precise and relevant legal information, aiding in case preparation and research. Scientific research could see improvements in literature review and hypothesis generation through more accurate and contextually aware information retrieval. The integration of domain-specific modules allows for a tailored approach that can be adapted to various fields, making the model a valuable tool in numerous professional and academic contexts. However, it is important to consider the limitations of the approach, including the need for extensive domain-specific datasets and the potential for overfitting if the training data is not sufficiently diverse.

The knowledge domain partitioning approach presents a promising strategy for enhancing the performance of large language models by allowing for more specialized domain knowledge concentration. The results of this study demonstrate the potential of the approach to significantly improve the accuracy and relevance of model responses within specific domains. The successful implementation of this strategy in the Llama model demonstrates its viability and applicability across various fields. However, addressing the limitations and challenges associated with the approach is crucial for its continued development and optimization. By building on the findings of this study, future research can further refine the approach, paving the way for more intelligent and specialized language models capable of delivering high-quality, contextually relevant information across diverse domains.

While the study demonstrates the benefits of knowledge domain partitioning, it also highlights several limitations that warrant further investigation. The reliance on large, high-quality domain-specific datasets may pose challenges in fields where such data is scarce or difficult to obtain. Additionally, the architectural complexity introduced by integrating domain-specific modules could limit the scalability and adaptability of the approach to other models or applications. The potential for overfitting, where the model becomes too specialized and loses its ability to generalize across domains, is another concern that must be addressed. These limitations suggest the need for ongoing research to refine and optimize the approach, ensuring it remains both effective and adaptable across different contexts.

Future research should focus on exploring methods to mitigate the limitations identified in this study, such as developing techniques to augment domain-specific datasets when data is limited. One potential avenue is the use of synthetic data generation to supplement real-world datasets, thereby enhancing the diversity and robustness of the training data. Additionally, research could explore more efficient ways to integrate domain-specific knowledge without significantly increasing the architectural complexity of the model. Investigating techniques such as transfer learning and few-shot learning could provide valuable insights into how to maintain a balance between specialization and generalization. Further studies could also examine the application of the knowledge domain partitioning approach to other types of language models, assessing its effectiveness and adaptability in different contexts.

VI. Conclusion

The key findings of this study reveal that the implementation of knowledge domain partitioning in the Llama model significantly enhances its performance metrics, such as accuracy, precision, recall, and F1-score, by enabling a more focused and specialized approach to domain-specific knowledge. The partitioning strategy, which involves segmenting the training

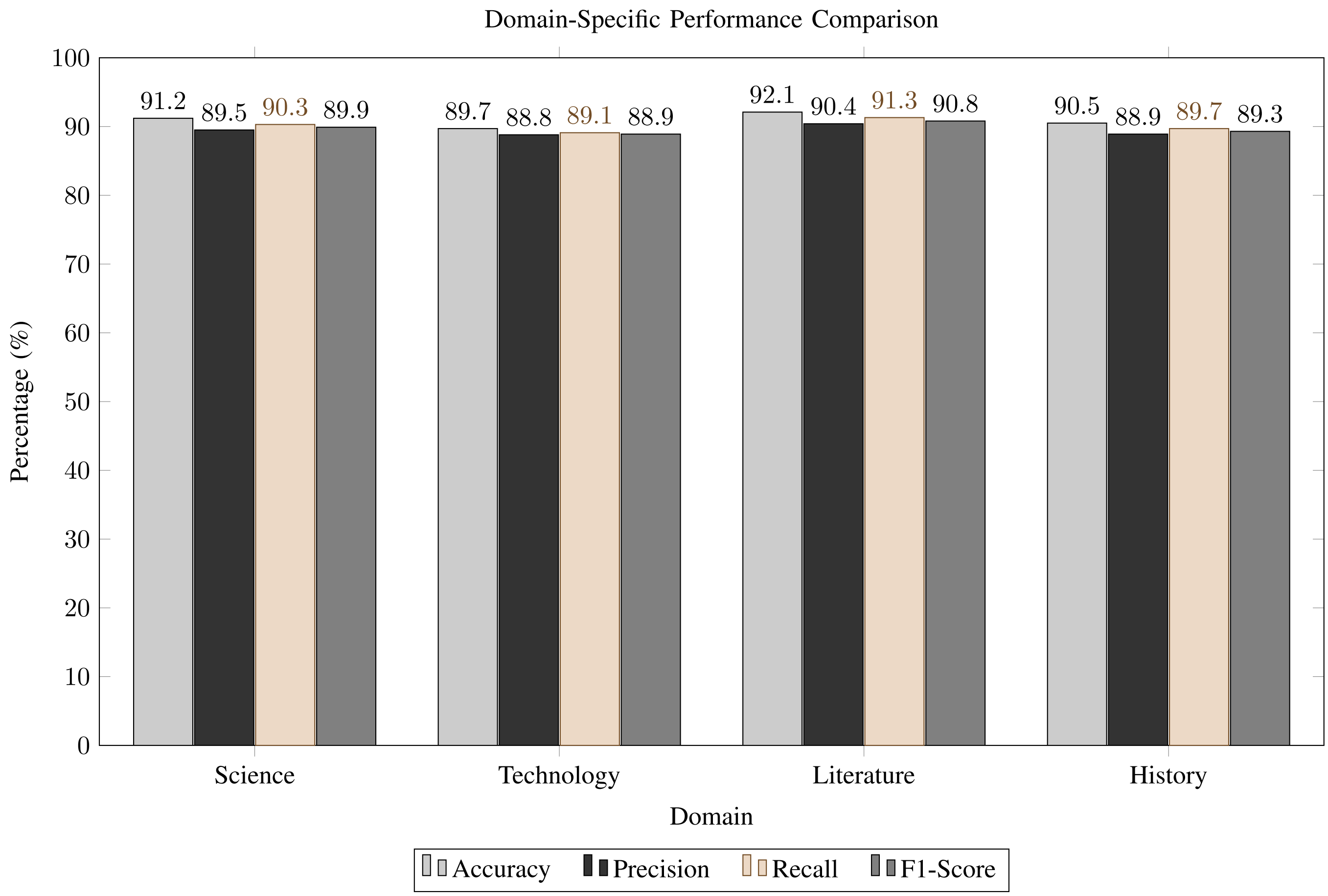

Fig. 2. Comparison of Domain-Specific Performance Metrics

data into distinct knowledge domains and modifying the model architecture to include domain-specific modules, proves to be highly effective in improving the contextual relevance and accuracy of responses. The significance of knowledge domain partitioning in large language models lies in its ability to address the inherent limitations of generalized knowledge representation, thereby providing a more robust and versatile tool for applications requiring specialized knowledge. By demonstrating substantial improvements in domain-specific performance, this research contributes valuable insights into the potential of domain-specific training strategies to enhance the capabilities of large language models. The findings demonstrate the importance of continued exploration and refinement of domain-specific methodologies, paving the way for the development of more intelligent and specialized language models capable of delivering high-quality, contextually relevant information across diverse fields. This study lays the groundwork for future research aimed at optimizing and extending the knowledge domain partitioning approach, ensuring its applicability and effectiveness in a wide range of professional and academic contexts. The successful implementation of this strategy in the Llama model serves as a testament to its viability, highlighting its potential to significantly advance the field of natural language processing.

References

[1] Y. Ge, W. Hua, K. Mei, J. Tan, S. Xu, Z. Li, Y. Zhang *et al.*, "Openagi: When llm meets domain experts," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[2] A. Sticha, "Utilizing large language models for question answering in task-oriented dialogues," 2023.

[3] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, no. 7972, pp. 172–180, 2023.

[4] B. M. Saiful, "Transfer learning for language model adaptation," 2023.

[5] L. Li, "Adapting pretrained vision-language models in medical domains," 2024.

[6] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[7] A. Fichtl, "Evaluating adapter-based knowledge-enhanced language models in the biomedical domain," 2024.

[8] J. Liuska, "Enhancing large language models for data analytics through domain-specific context creation," 2024.

[9] E. Hajihashemi Varnousfaderani, "Challenges and insights in semantic search using language models," 2023.

[10] Z. Chen, H. Mao, H. Li, W. Jin, H. Wen, X. Wei, S. Wang, D. Yin, W. Fan, H. Liu *et al.*, "Exploring the potential of large language models (llms) in learning on graphs," *ACM SIGKDD Explorations Newsletter*, vol. 25, no. 2, pp. 42–61, 2024.

[11] A. Caballero Hinojosa, "Exploring the power of large language models: News intention detection using adaptive learning prompting," 2023.

[12] D. Christofidellis, G. Giannone, J. Born, O. Winther, T. Laino, and M. Manica, "Unifying molecular and textual representations via multi-task language modelling," in *International Conference on Machine Learning*. PMLR, 2023, pp. 6140–6157.

[13] M. A. Lowe, "Ocr2seq: A novel multi-modal data augmentation pipeline for weak supervision," 2023.

[14] B. D. Davis, "Exploratory search with archetype-based language models," 2021.

[15] J. H. Kim and H. R. Kim, "Cross-domain knowledge transfer without retraining to facilitating seamless knowledge application in large language models," 2024.

[16] E. A. Kowalczyk, M. Nowakowski, and Z. Brzezińska, "Designing incremental knowledge enrichment in generative pre-trained transformers," 2024.

[17] T. Li, "Explore the in-context learning capability of large language models," 2024.

[18] T. R. McIntosh, T. Liu, T. Susnjak, P. Watters, A. Ng, and M. N. Halgamuge, "A culturally sensitive test to evaluate nuanced gpt hallucination," *IEEE Transactions on Artificial Intelligence*, 2023.

[19] L. Secchi *et al.*, "Knowledge graphs and large language models for intelligent applications in the tourism domain," 2024.

[20] M. Klettner, "Augmenting knowledge-based conversational search systems with large language models," 2024.

[21] K. Qiu, "Personal intelligent assistant based on large language model: Personalized knowledge extraction and query answering using local data and large language model," 2024.

[22] D. McDonald, R. Papadopoulos, and L. Benningfield, "Reducing llm hallucination using knowledge distillation: A case study with mistral large and mmlu benchmark," *Authorea Preprints*, 2024.

[23] Y. Chen, "An intelligent question-answering system for course learning based on knowledge graph," 2024.

[24] P. M. Whitehead, "Multilingual extractive question answering with conflibert for political and social science studies," 2023.

[25] S. Sakhinana and V. Runkana, "Cross-modal learning for chemistry property prediction: Large language models meet graph machine learning," in *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*, 2023.

[26] K. Kobs, "Model-agnostic deep learning with domain knowledge," 2024.

[27] D. Lukovnikov, "Deep learning methods for semantic parsing and question answering over knowledge graphs," 2022.

[28] B. Guo, H. Wang, Y. Ding, W. Wu, S. Hao, Y. Sun, and Z. Yu, "Conditional text generation for harmonious human-machine interaction," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 2, pp. 1–50, 2021.

[29] L. A. K. Pigott-Dix, "Automating the annotation of data through machine learning and semantic technologies," 2023.

[30] C. Campàs Gené, "Fine tuning transformer models for domain specific feature extraction," 2023.

[31] A. Ghosh, A. Acharya, R. Jain, S. Saha, A. Chadha, and S. Sinha, "Clipsyntel: clip and llm synergy for multimodal question summarization in healthcare," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, 2024, pp. 22 031–22 039.