Neural Symbol and Large Language Models in Reasoning and Problem Solving: A Survey

www.surveyx.cn

Abstract

The interdisciplinary domain where neural networks, large language models (LLMs), and symbolic reasoning converge represents a pivotal frontier in artificial intelligence (AI), enhancing problem-solving capabilities in logical and mathematical contexts. This survey explores the integration of these technologies, highlighting their synergistic potential to address complex tasks requiring deep learning and structured logical processing. Neural networks' pattern recognition abilities complement LLMs like GPT-3.5 and GPT-4, which excel in language comprehension and generation, while symbolic reasoning offers the precision necessary for high-level cognitive functions. This synthesis yields numerous advantages, including improved adaptability, interpretability, and efficiency, crucial for applications spanning healthcare, education, autonomous systems, and creative industries. Frameworks such as GraphAgent-Reasoner and RLTF exemplify innovative approaches that enhance scalability and robustness, facilitating the seamless combination of structured logic with dynamic learning environments. Despite significant progress, challenges persist in optimizing computational efficiency, data quality, model robustness, and scalability. Addressing these is essential for maximizing the potential of neural-symbolic systems and ensuring responsible deployment across various fields. The exploration of new learning paradigms, such as hybrid frameworks and continual learning techniques, promises further advancements, enabling systems to perform complex reasoning tasks with greater precision and adaptability. As the field evolves, the formalization of neural-symbolic design patterns provides a clearer framework for future AI advancements, expanding the problem-solving repertoire and paving the way for more interpretable, trustworthy, and efficient AI systems.

1 Introduction

1.1 Interdisciplinary Domain Overview

The interdisciplinary domain where neural networks, large language models (LLMs), and symbolic reasoning converge marks a pivotal area in artificial intelligence research, enhancing problem-solving capabilities in logical and mathematical contexts. Neural networks excel in pattern recognition, while LLMs, such as GPT-3.5 and GPT-4, demonstrate remarkable proficiency in language comprehension and generation, facilitating advanced reasoning processes. This synergy is critical for addressing complex tasks, such as the detection of AI-generated text in academia, where concerns about integrity and plagiarism are increasingly relevant [1, 2].

The integration of LLMs with symbolic reasoning has improved AI performance in low-data environments, broadening AI applicability across diverse fields. In robotics, this fusion enhances task planning, showcasing LLMs' potential to support robotic manipulation and problem-solving in real-world scenarios [3]. Moreover, in autonomous driving technologies, the interplay between AI and LLMs underscores the importance of decision-making processes, illustrating LLMs' role in facilitating complex communication systems [4].

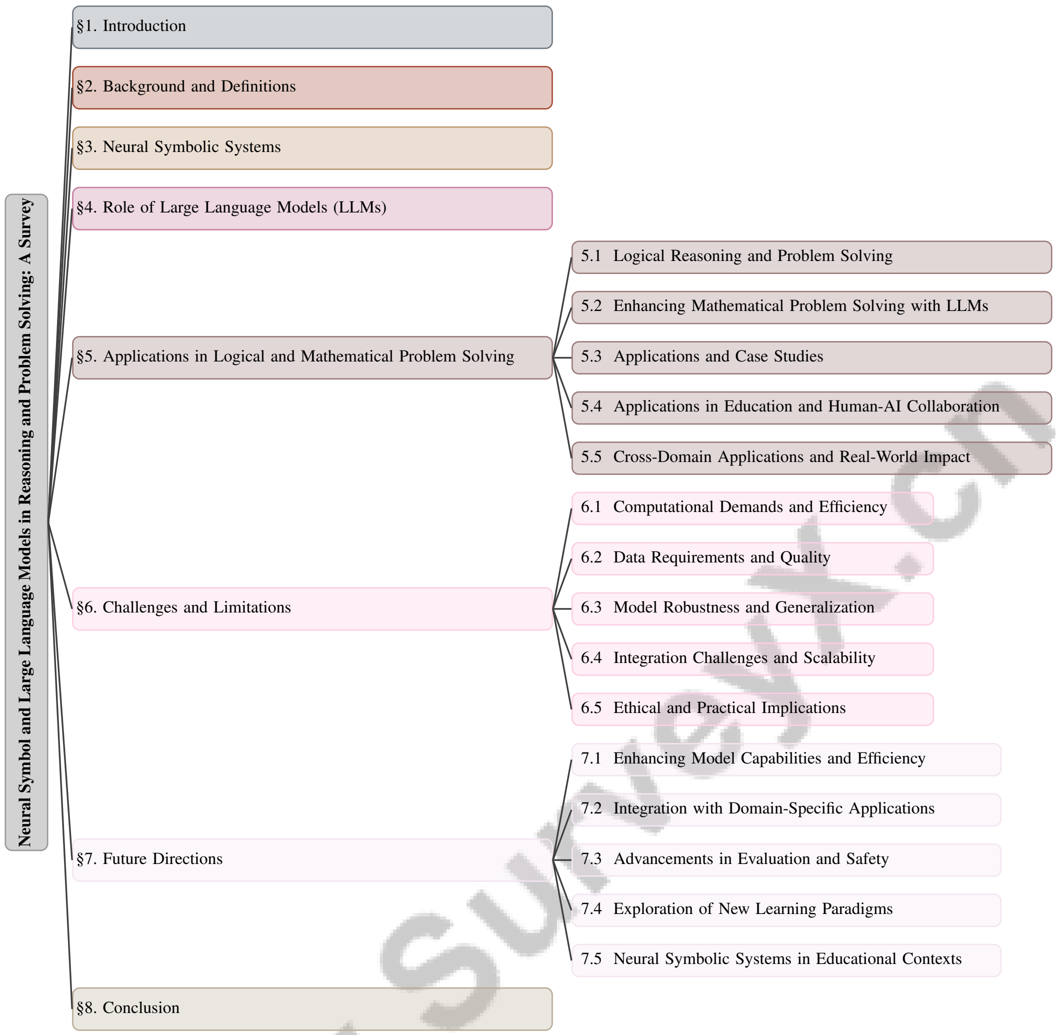


Figure 1: chapter structure

LLMs significantly contribute to business process management and educational applications, highlighting their transformative potential across sectors. The integration of structured emergency knowledge with LLMs enhances decision-making during crises, providing evidence-based support that mitigates cognitive limitations. For instance, the E-KELL system employs knowledge graphs to guide LLMs in reasoning, improving the comprehensibility and accuracy of emergency responses. This interdisciplinary approach not only demonstrates AI's potential in emergency management but also emphasizes the necessity of combining machine intelligence with structured knowledge for reliable decision support in high-stakes situations [5, 6].

Furthermore, the intersection of LLMs with symbolic reasoning and neural networks addresses the demand for principled knowledge representation and reasoning mechanisms, ensuring AI systems' interpretability and accountability [7]. This approach is vital for overcoming challenges related to LLMs generating incorrect answers due to their reliance on implicit knowledge, necessitating enhanced inference control and explainability [8]. As the field evolves, the dynamic interaction between humans and LLMs presents both challenges and opportunities, warranting careful consideration of ethical implications and collaborative advancements in AI technologies [9].

The integration of LLMs with formal methods and wavelet transforms further exemplifies the interdisciplinary nature of this domain, merging natural language processing with advanced data structures to bolster AI's reasoning capabilities and model efficiency. Incorporating LLMs into interpretable machine learning is essential for redefining interpretability, facilitating nuanced and natural language explanations while supporting the development of smaller, cost-effective models.

This integration enables researchers to analyze complex datasets and produce interactive explanations, enhancing understanding and evaluation of machine learning systems while addressing challenges like hallucinated outputs and high computational demands [10, 11]. Additionally, the synergy between AI and education underscores LLMs' potential to transform pedagogical approaches, expanding the impact of AI technologies across various domains.

1.2 Significance of Integration

The integration of neural networks, large language models (LLMs), and symbolic reasoning signifies a transformative leap in artificial intelligence (AI), enabling systems to address complex, knowledge-intensive tasks with enhanced precision and contextual relevance. In healthcare, LLMs applied to clinical contexts illustrate the potential for improved decision support systems, positively influencing clinical outcomes and reliability in emergency management scenarios [9]. This integration is equally vital in software development and neural architecture search, where LLMs, combined with optimization algorithms, have led to significant efficiency gains and complexity reductions [3].

The incorporation of LLMs into frameworks like Monte Carlo Tree Search (MCTS) exemplifies their ability to enhance decision-making in mathematical contexts, providing a robust foundation for solving complex problems [12]. Similarly, the development of LLM-based agents, such as Richelieu, which integrates social reasoning, memory management, and strategic planning, demonstrates LLMs' potential to excel in dynamic environments like the Diplomacy game [4].

In multilingual contexts, methodologies like MindMerger aim to integrate LLMs with external multilingual capabilities, enhancing reasoning performance in low-resource languages and broadening AI's global applicability [13]. The combination of LLMs with symbolic reasoning also addresses hallucination challenges by incorporating knowledge graphs, which improve the reliability and accuracy of AI outputs across various applications [14].

Moreover, the iterative process of concept injection and refinement within LLMs seeks to enhance the creativity and adaptability of outputs, aligning them more closely with human evaluators' expectations [7]. Collectively, these integrations underscore the critical importance of combining neural networks, LLMs, and symbolic reasoning in expanding AI's problem-solving capabilities and fostering effective human-AI collaboration across diverse fields.

1.3 Structure of the Survey

The survey is structured into key sections to thoroughly investigate the intersection of neural networks, large language models (LLMs), and symbolic reasoning within artificial intelligence (AI). The introduction provides an overview of the interdisciplinary domain, highlighting the significance of integrating these technologies to enhance AI's problem-solving capacities in logical and mathematical contexts. The background section elucidates core concepts, establishing a foundational understanding of neural networks, LLMs, and symbolic reasoning's roles in AI systems.

Subsequently, the survey explores neural symbolic systems, discussing the integration of neural networks with symbolic reasoning frameworks. This section examines the advantages and challenges of such integration, offering insights into existing frameworks that facilitate this synergy. The role of LLMs is highlighted, emphasizing their capabilities in language understanding and generation, along with their contributions to knowledge enhancement in AI systems.

The survey further investigates applications in logical and mathematical problem-solving, presenting case studies and examples of neural symbolic systems and LLMs across various domains. Discussions encompass strategies for enhancing mathematical problem-solving through innovative tools like Mathemyths, which utilizes child-AI co-creative storytelling to teach mathematical language. Additionally, the applications of LLMs in education are explored, including automatic deductive coding for discourse analysis and advanced plagiarism detection methods that bolster academic integrity. The collaboration between humans and AI is also examined, showcasing LLMs' potential to assist in evaluating scientific ideas and addressing mathematical reasoning challenges. Finally, the real-world impact of these technologies is analyzed, demonstrating their effectiveness in educational contexts and their role in shaping future research methodologies [14, 15, 2, 16, 11].

Challenges and limitations are addressed in the subsequent section, focusing on computational demands, data requirements, model robustness, integration challenges, and ethical implications. The survey concludes with a discussion on future directions, exploring potential advancements in model capabilities, domain-specific applications, evaluation techniques, safety measures, and new learning paradigms within AI.

Throughout the survey, references to pertinent research papers underpin our discussions, establishing a scholarly narrative. Notably, one significant study explores LLMs' role in assessing scientific ideas using a curated benchmark dataset of nearly four thousand manuscripts, illustrating the potential of LLM representations in quantifying idea merit and suggesting a promising direction for automating idea assessment. Additionally, we reference a study on Retrieval-Augmented Generation (RAG), examining its limitations in enhancing LLM reasoning capabilities. This investigation reveals that while RAG can introduce new knowledge and mitigate hallucinations, its effectiveness is limited, particularly in facilitating deeper reasoning processes. Together, these references ensure a comprehensive and well-supported narrative that emphasizes the importance of rigorous evaluation and innovation in academic research [17, 11].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts of Neural Networks and LLMs

Neural networks and large language models (LLMs) are pivotal in advancing artificial intelligence (AI), each offering distinct capabilities that drive innovation across multiple domains. Inspired by biological systems, neural networks excel in pattern recognition and feature extraction from complex datasets, facilitating tasks such as image and speech recognition through their layered neuron structures [14]. Their adaptability allows simulation of Turing-complete systems, making them versatile for a wide array of AI applications.

LLMs, including models like GPT-3.5 and GPT-4, have transformed natural language processing with their transformer-based architectures, enabling fluent text generation and comprehension [12]. Despite their advancements, challenges in mathematical problem-solving and logical reasoning persist, requiring enhanced reasoning strategies. The stochastic nature of LLMs can result in confabulations, impacting reliability in sensitive areas such as healthcare [9]. This unpredictability underscores the need for robust benchmarks, which are still evolving [18].

Optimizing LLMs is critical to enhance inference efficiency given their computational demands. Techniques like pre-training and fine-tuning are vital for improving performance, allowing adaptation to specific tasks and domains. Methods such as Reasoning-Aware Self-Consistency (RASC) enhance sampling efficiency by evaluating reasoning paths [12]. However, LLMs struggle with inductive learning over complex rule sets due to their limited symbolic operation capabilities [7].

In educational contexts, LLMs support dialog-based tutoring systems, though challenges in pedagogical strategy development and dataset costs persist. Their flexibility aids in qualitative analysis, yet inaccuracies necessitate ongoing improvements [19]. The potential for encoded reasoning, which may conceal reasoning steps, raises concerns about output faithfulness.

LLMs encapsulate extensive factual knowledge in their parameters, but conflicts with contextual information can arise. They face difficulties with memory and incorporating new information, especially in diverse educational domains. Hypotheses suggest LLMs can adjust outputs based on inputs, paralleling human memory mechanisms, indicating adaptability and human-like cognitive processing. This adaptability is evident in their ability to transform complex ideas into quantifiable features, enhancing predictive analytics [20, 11, 21].

LLMs also face challenges in SQL querying due to undefined schemas and extracting structured data from unstructured text. Their graph reasoning scalability and accuracy are limited by graph complexity and long-form data processing constraints. Recent advancements like the GraphAgent-Reasoner framework address these challenges through multi-agent collaboration, enhancing accuracy and scalability for larger graphs. Frameworks such as Graph-ToolFormer aim to improve graph reasoning via prompt augmentation and API integration, though difficulties remain in advanced graph tasks [22, 23, 5, 16, 24]. Efficient approaches are needed due to the extensive training resources required for optimal LLM performance.

Neural networks and LLMs are transformative in AI, driving progress across diverse fields. Their integration and optimization are crucial for overcoming limitations and enhancing capabilities in both language and non-language tasks. Addressing vulnerabilities related to security, privacy, and ethical implications remains essential as these technologies evolve [14].

2.2 Symbolic Reasoning in AI

Symbolic reasoning is a fundamental component of artificial intelligence (AI), leveraging formal logic and symbol manipulation for tasks like problem-solving, decision-making, and deriving conclusions from established premises. This approach is crucial for high-level cognitive functions, enabling AI systems to emulate human-like reasoning through structured frameworks [25]. By employing first-order logic, symbolic reasoning enhances decision-making processes and ensures logical consistency across various applications [24].

Despite advancements in neural networks and LLMs, symbolic reasoning remains indispensable due to its precision and ability to handle explicit knowledge representation. While neural architectures like Transformers can simulate Turing-complete systems, they often lack the inherent ability to perform symbolic computations without external memory [26]. This limitation highlights the complementary role of symbolic reasoning in augmenting LLMs' reasoning capabilities, particularly in precision-critical domains.

Integrating symbolic reasoning with LLMs addresses issues such as generating plausible yet incorrect outputs, known as hallucinations [27]. These arise from the probabilistic nature of LLMs and their limitations in complex computations or real-time information access. Incorporating symbolic reasoning can mitigate these issues by enhancing models' ability to identify knowledge gaps and reduce biases, leading to more reliable outputs [28].

In mathematical reasoning, symbolic reasoning is vital due to the complexity involved in scaling and capacity [25]. Symbolic methods provide the precision necessary for accurately solving mathematical problems, as evidenced by benchmarks designed to simulate real-world educational challenges [29]. These benchmarks not only evaluate LLM proficiency in mathematical reasoning but also highlight the importance of symbolic reasoning in enhancing AI's cognitive modeling capabilities [30].

Challenges in evaluating AI systems using traditional methods reveal inadequacies in capturing nuanced attributes in AI and NLP tasks, necessitating symbolic reasoning for a comprehensive evaluation framework [18]. The dynamic nature of real-world data and the need for LLMs to adapt while retaining learned information further emphasize symbolic reasoning's role in ensuring AI systems remain robust and adaptable.

3 Neural Symbolic Systems

Neural symbolic systems represent a significant evolution in artificial intelligence (AI), merging neural networks and symbolic reasoning to enhance reasoning capabilities and problem-solving efficiency. This section explores the integration of these paradigms, illustrating how their combination improves AI's adaptability and interpretability across various applications. Table 1 presents a comparative analysis of the integration methods, advantages, and frameworks of neural networks and symbolic reasoning, illustrating their combined impact on AI capabilities. As depicted in Figure 2, the hierarchical structure of neural symbolic systems highlights the integration of neural networks and symbolic reasoning, emphasizing their advantages and the various frameworks that facilitate this integration. This visual representation underscores how this integration enhances AI's problem-solving capabilities, adaptability, and interpretability, particularly in fields such as diplomacy, education, and healthcare.

3.1 Integration of Neural Networks and Symbolic Reasoning

The fusion of neural networks with symbolic reasoning systems marks a substantial advancement in AI, marrying the learning adaptability of neural architectures with the precision of symbolic logic. This synergy enhances AI's ability to tackle complex reasoning tasks, evidenced by techniques such as automatic deductive coding and Retrieval-Augmented Generation (RAG), which improve reasoning processes [14, 2, 17, 31, 21]. As illustrated in Figure 3, this figure highlights the integration

AI-generated, for reference only.

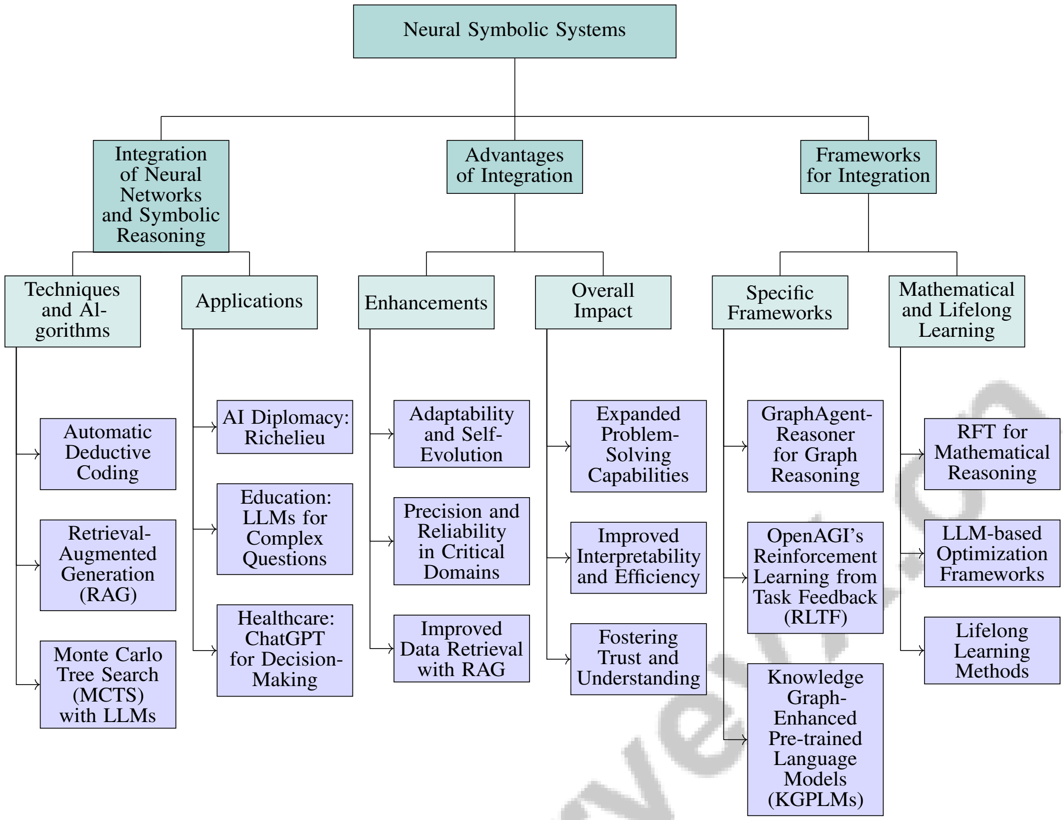


Figure 2: This figure illustrates the hierarchical structure of neural symbolic systems, highlighting the integration of neural networks and symbolic reasoning, their advantages, and various frameworks that facilitate this integration. The integration enhances AI's problem-solving capabilities, adaptability, and interpretability across applications like diplomacy, education, and healthcare.

of neural networks and symbolic reasoning, showcasing enhancement techniques, applications, and key benefits that contribute to AI's evolution.

Monte Carlo Tree Search (MCTS) algorithms, when integrated with large language models (LLMs), exemplify the potential of combining probabilistic search with sophisticated language capabilities, facilitating robust problem-solving [12]. In AI diplomacy, LLM-based agents like Richelieu utilize memory management and strategic planning to navigate complex negotiations, showcasing the integration's effectiveness in dynamic environments [4]. The use of structured protocols enhances AI communication, bridging the gap between machine and human language [7]. In education, LLMs generate complex questions, demonstrating the integration's capacity to support learning through sophisticated reasoning [19]. In healthcare, applications like ChatGPT improve decision-making and planning, illustrating the practical benefits of neural-symbolic systems in specialized domains [9].

Overall, this integration is pivotal in AI's evolution, enhancing problem-solving capabilities and facilitating the development of interpretable and efficient systems applicable in fields such as plagiarism detection, idea assessment, and predictive analytics [2, 31, 1, 11, 21].

3.2 Advantages of Integration

Integrating neural networks with symbolic reasoning systems offers significant enhancements in AI capabilities. Systems like Richelieu demonstrate adaptability through self-evolution, crucial for AI in dynamic environments [4]. In education, this integration generates challenging questions at scale, supporting personalized learning [19]. The integration also improves AI precision and reliability, essential in critical domains like healthcare and emergency management, where frameworks such as E-KELL enhance decision accuracy [2, 32, 6, 11, 21].

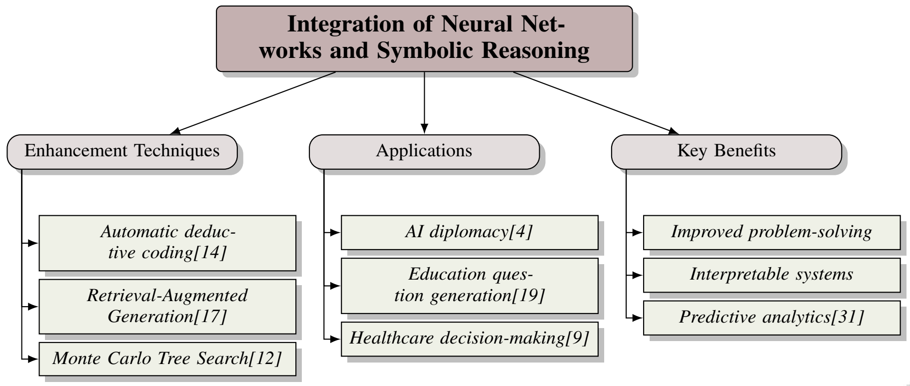


Figure 3: This figure illustrates the integration of neural networks and symbolic reasoning, highlighting enhancement techniques, applications, and key benefits in AI's evolution.

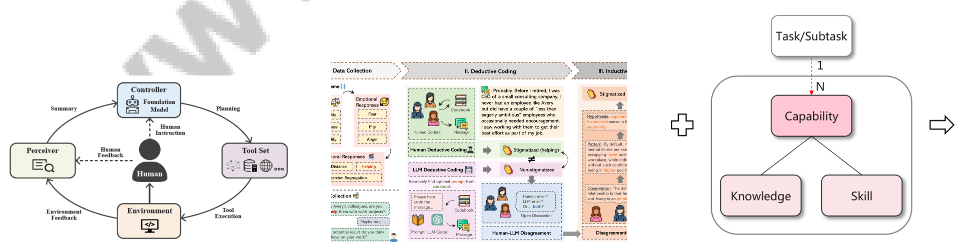
LLMs enhance data retrieval processes, leveraging techniques like RAG to incorporate domain-specific knowledge, beneficial in intricate data manipulation applications [2, 17, 31, 33, 11]. Overall, this integration expands AI's problem-solving capabilities, improving interpretability, efficiency, and adaptability, fostering trust and understanding among users [34, 11, 35, 2].

3.3 Frameworks for Integration

Various frameworks facilitate the integration of neural networks and symbolic reasoning, enhancing AI capabilities. The GraphAgent-Reasoner framework advances graph reasoning tasks, enabling LLMs to handle complex data structures efficiently [24]. OpenAGI's Reinforcement Learning from Task Feedback (RLTF) mechanism emphasizes feedback-driven learning for AI model refinement [36]. Knowledge graph-enhanced pre-trained language models (KGPLMs) integrate structured knowledge into neural models at various stages [37].

In mathematical reasoning, RFT enhances accuracy by utilizing structured reasoning paths [38]. LLM-based optimization frameworks demonstrate versatility in enhancing AI performance [39]. Lifelong learning methods ensure LLMs remain adaptive, handling evolving tasks effectively [40].

These frameworks illustrate diverse approaches to integrating neural networks and symbolic reasoning, emphasizing structured knowledge and feedback mechanisms in enhancing AI's problem-solving capabilities. As these frameworks evolve, they enhance AI systems' ability to address complex tasks, improving accuracy and transparency while preserving critical human expertise [14, 34, 2, 21].



(a) Human-robot interaction framework[41]

(b) Human-LLM Deductive Coding: A Framework for Identifying Stigmatization in Emotional Responses[42]

(c) The image represents a hierarchical structure of knowledge and skills related to a task or subtask.[43]

Figure 4: Examples of Frameworks for Integration

As depicted in Figure 4, the exploration of neural symbolic systems is a promising area in AI, bridging symbolic reasoning and neural networks. Frameworks such as the "Human-robot interaction framework" illustrate comprehensive systems where human inputs and robotic responses are facilitated by advanced model-based planning. The "Human-LLM Deductive Coding" framework identifies stigmatization in emotional responses, leveraging LLMs in nuanced applications. The hierarchical structure of knowledge and skills emphasizes the interconnectedness of capability, knowledge, and skill. These examples underscore neural symbolic systems' versatility in applications ranging from human-robot interaction to refining AI's emotional intelligence [41, 42, 43].

Feature	Integration of Neural Networks and Symbolic Reasoning	Advantages of Integration	Frameworks for Integration
Integration Method	Fusion OF Systems	Enhanced Capabilities	Structured Knowledge
Key Application	Complex Reasoning	Dynamic Environments	Problem-solving
Framework Example	Mcts With Llms	Richelieu Agents	Graphagent-Reasoner

Table 1: This table provides a comparative overview of the integration of neural networks and symbolic reasoning, highlighting the methods, key applications, and frameworks involved. It emphasizes the benefits of such integration, including enhanced capabilities and problem-solving efficiency, through examples of specific frameworks employed.

4 Role of Large Language Models (LLMs)

4.1 Capabilities in Language Understanding and Generation

Large Language Models (LLMs) like GPT-3.5 and GPT-4 represent a significant leap in natural language processing (NLP) by mastering complex linguistic structures and generating contextually relevant text. Built on transformer architectures, these models enhance decision-making with nuanced insights. Techniques such as Graph-Constrained Reasoning (GCR) ensure language generation fidelity by anchoring reasoning paths in knowledge graphs [23]. LLMs also excel in solving complex mathematical problems, as illustrated by the MCT Self-Refine algorithm, which improves success rates across datasets, showcasing their ability to tackle intricate reasoning tasks [12].

Their adaptability is further evidenced by SQL querying capabilities, facilitating structured data retrieval from unstructured sources [44]. The LLM Cascade strategy with Multi-Objective Optimization supports decision-making by considering multiple objectives beyond performance metrics [45]. In creative language generation, models like Llama 3.1 INSTRUCT generate innovative concepts evaluated against creativity metrics, highlighting their potential in language tasks [8]. This creative ability is enhanced by iterative refinement of theories based on feedback, improving logical reasoning processes [46].

In education, LLMs generate questions requiring deeper reasoning and skill application, merging human intuition with computational power [19]. This synergy supports complex educational tasks. The hierarchical categorization of capabilities in language understanding and generation is illustrated in Figure 5, highlighting advanced reasoning techniques, creative and structured outputs, and tool integration and adaptation as primary categories, supported by specific methodologies and innovations. LLMs' proficiency in language understanding and generation is characterized by their ability to process complex linguistic structures, adapt to diverse contexts, and learn from iterative feedback. They are transformative tools in NLP and human-computer interaction, integrating external tools to enhance efficiency and accuracy in complex tasks. This evolution broadens their applicability across fields like medicine and psychology, addressing challenges in user intent understanding and dynamic plan adjustment. Through innovative fine-tuning and in-context learning techniques, LLMs adapt responses more effectively, improving evaluation mechanisms that yield deeper insights into performance. Consequently, LLMs are transitioning from passive tools to potential creators, redefining their role in technology and interaction paradigms [47, 41].

4.2 Roles of LLMs in Knowledge Enhancement

LLMs significantly enhance knowledge within AI systems by integrating vast information and supporting complex decision-making processes. Their advanced natural language processing capabilities improve human-AI collaboration quality, with understanding determinants in LLM-assisted decision-making leading to better interfaces and user experiences [48]. In mathematical reasoning, LLMs push the boundaries of AI-driven reasoning by evaluating performance on benchmarks involving open-ended questions from platforms like Math Stack Exchange [16]. The application of GFlowNet fine-tuning allows LLMs to generate diverse and accurate solutions in mathematical tasks, showcasing adaptability and precision in complex problem-solving [49].

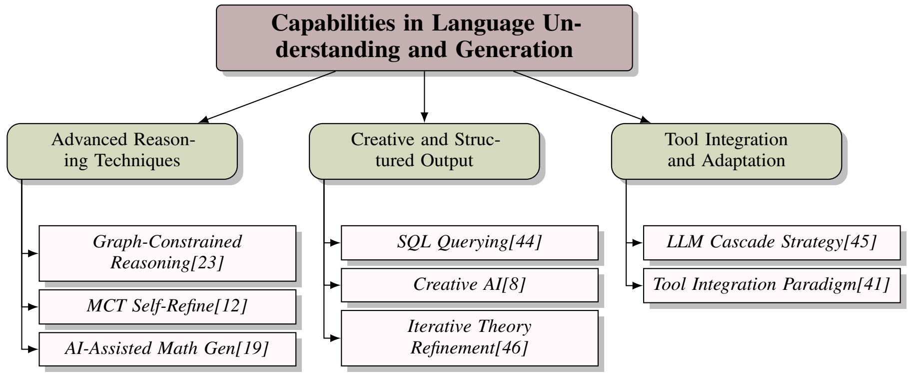


Figure 5: This figure illustrates the hierarchical categorization of capabilities in language understanding and generation, highlighting advanced reasoning techniques, creative and structured outputs, and tool integration and adaptation as primary categories, supported by specific methodologies and innovations.

In education, LLMs are invaluable, with generated tips rivaling expert-created content in fields like quantum computing, reducing educators' workloads while maintaining educational quality [50]. LLMs also enhance knowledge through efficient and cost-effective deployment solutions, achieving significant cost savings and memory efficiency across various quantization bit-widths, ensuring state-of-the-art performance while managing resource constraints [51]. Incorporating multi-objective considerations into LLM cascades enhances performance and reduces privacy risks, crucial for real-world applications [45].

The expressiveness of LLMs is enhanced by techniques like SaySelf, which improves models' confidence expression, leading to better interaction and performance in uncertain scenarios [52]. This enhancement is crucial for applications requiring nuanced communication and decision-making.

5 Applications in Logical and Mathematical Problem Solving

This section explores the applications of Large Language Models (LLMs) in logical and mathematical problem-solving, emphasizing their foundational role in addressing complex challenges through systematic reasoning and structured methodologies. LLMs' contributions are highlighted through practical applications and case studies across various contexts, demonstrating their effectiveness in enhancing logical reasoning and problem-solving capabilities.

5.1 Logical Reasoning and Problem Solving

LLMs exhibit significant potential in logical reasoning and problem-solving across diverse applications. In robotics, LLMs enhance task-oriented success in dynamic environments, outperforming traditional methods by improving goal condition recall and task success in tasks like 'catch' and 'clean the top of the cabinet' [53, 54]. Their capabilities extend to ontology learning tasks, including term typing and taxonomy discovery, as demonstrated by experiments with Graph-ToolFormer in graph reasoning tasks such as bibliographic topic inference [55, 22]. LLMs also enhance multi-robot task planning, showcasing their collaborative potential [56].

In standardized testing, LLMs' potential is evaluated using GMAT questions, highlighting their role in structured problem-solving [57]. The E-KELL system exemplifies LLMs' logical reasoning in emergency scenarios like hazardous chemical leaks [6]. Innovative pre-training techniques, such as wavelet transforms, significantly enhance performance, achieving faster results across data modalities [58]. LLMs also demonstrate effective inductive learning capabilities, rivaling traditional ILP systems in noise handling [46].

LLMs' advantages include low cost, rapid execution, and the exploration of vast parameter spaces without ethical concerns, positioning them as valuable tools in advancing logical reasoning and problem-solving across fields like automated ontology learning and emergency decision-making [59].

5.2 Enhancing Mathematical Problem Solving with LLMs

LLMs have significantly advanced mathematical problem-solving through innovative methodologies. The AutoMathCritique framework automates critique data collection, refining LLMs' reasoning processes and improving problem-solving strategies [60]. The RFT technique integrates diverse reasoning paths into training data, enhancing mathematical reasoning performance [38]. Knowledge Graphs (KGs) alongside LLMs enhance educational question-answering systems, deepening understanding of mathematical concepts [5].

Psychometrics-assisted benchmarking, using datasets like Patch, provides a robust framework for evaluating LLMs' proficiency across mathematical domains [61]. Fine-tuning on datasets like MMIQC has demonstrated substantial accuracy improvements, underscoring the importance of targeted training [62]. The Reasoning-Token Economies (RTE) dataset emphasizes budget-aware evaluations, contributing to effective and resource-conscious LLMs [26].

Progressive Rectification Prompting (PRP) achieves remarkable accuracy across math word problem datasets, enhancing LLMs' reasoning capabilities [63]. Fine-tuning with GFlowNet generates diverse yet accurate derivations, vital for developing adaptable LLMs [49]. The MCTSr algorithm enhances smaller parameter models' mathematical reasoning, democratizing access to high-performance LLMs [12]. An AI-assisted pipeline for generating difficult math problems illustrates LLMs' broader applicability, revolutionizing problem-solving across various domains [19].

5.3 Applications and Case Studies

LLMs and neural-symbolic systems are applied across domains, enhancing reasoning and problem-solving capabilities. Cooperative strategic planning improves performance on multi-choice reasoning benchmarks like LogiQA and BBH, emphasizing LLMs' effectiveness in complex reasoning tasks [64]. Ethical considerations are critical in understanding LLM ethics across sectors, with case studies highlighting the need for responsible AI practices [65].

DPrompt tuning enhances reasoning capabilities by strategically positioning queries, optimizing LLM performance in information retrieval and data processing [17]. These case studies underscore LLMs' transformative potential in fields like education, ethics, strategic planning, and information retrieval. Ongoing experimentation and adaptation are essential to leverage advanced AI systems effectively, particularly in contexts like plagiarism detection and predictive analytics [2, 21].

As illustrated in Figure 6, which encapsulates the key applications and case studies of large language models (LLMs), the integration of advanced computational frameworks and synthetic systems demonstrates potential in logical and mathematical problem-solving. The figure highlights frameworks such as CoPlanner and DPrompt tuning for LLM enhancements, alongside surveys on LLM ethics and responsible AI practices. Additionally, it showcases educational innovations like the Synthetic Tutoring System for Reading Comprehension, which bolsters reading skills through structured worksheets and synthetic dialogues, simulating a tutoring environment that adapts to learners' needs. The Open Entity Discovery Framework efficiently manages and organizes data, enhancing entity discovery and categorization through techniques like sentence embedding. These examples collectively underscore technology's role in improving educational outcomes and data management processes [66, 5].

5.4 Applications in Education and Human-AI Collaboration

Integrating neural-symbolic systems and LLMs into educational contexts and human-AI collaboration enhances learning outcomes and fosters innovative teaching methodologies. The CHALET method exemplifies human-LLM collaboration in qualitative research, facilitating deeper insights and comprehensive analyses in educational settings [42]. By leveraging LLMs' strengths in processing large datasets, educators can enhance research capabilities, leading to informed pedagogical strategies and improved educational outcomes.

'Prompt Problems' in education enhance student learning and engagement, particularly in programming tasks, by encouraging active learning and fostering critical thinking and problem-solving abilities [67]. These applications highlight the potential of neural-symbolic systems and LLMs to transform learning environments, facilitating innovative human-AI collaboration and enabling personalized and effective educational experiences.

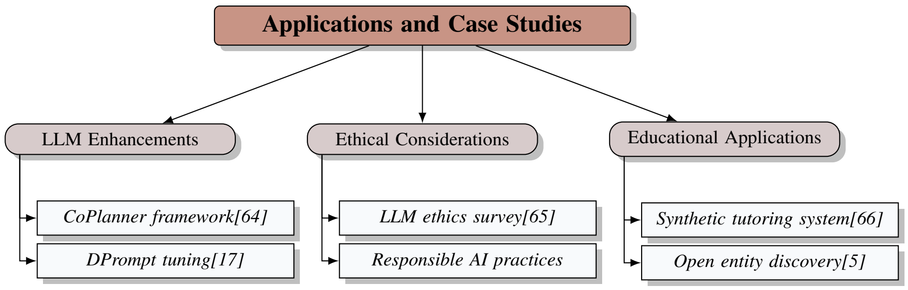


Figure 6: This figure illustrates the key applications and case studies of large language models (LLMs), focusing on enhancements in reasoning capabilities, ethical considerations, and educational applications. It highlights frameworks like CoPlanner and DPrompt tuning for LLM enhancements, surveys on LLM ethics and responsible AI practices, and educational innovations such as synthetic tutoring systems and open entity discovery.

Advancements in chatbot implementations in specialized domains like medicine and psychology have prompted discussions on robust evaluation mechanisms. As LLMs generate sophisticated content, ethical concerns regarding academic integrity arise, necessitating solutions for detecting AI-generated text [47, 1, 2]. By integrating AI into educational practices, these systems enhance teaching and learning and prepare students for the complexities of the modern technological landscape.

5.5 Cross-Domain Applications and Real-World Impact

Integrating neural-symbolic systems and LLMs transforms sectors by enhancing decision-making processes and operational efficiencies. This synergy enables LLMs to leverage extensive domain knowledge for intelligent modeling and utilize optimization algorithms to refine their architecture and output quality. Industries can better assess scientific ideas, automate evaluation, and generate interpretable insights, revolutionizing workflows and improving human-computer interactions [10, 3, 11, 41]. In healthcare, LLMs improve diagnostic accuracy and treatment planning, facilitating access to vast medical data.

In business process management, LLMs optimize workflows and enhance operational efficiency, particularly in multilingual contexts, supporting seamless communication and data processing [13]. This capability is crucial for global enterprises requiring effective coordination across diverse linguistic and cultural environments.

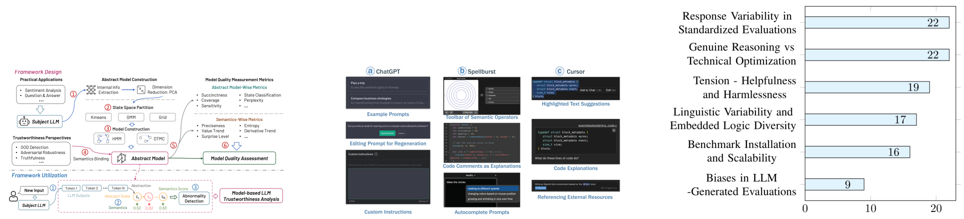
The educational sector benefits from LLMs, supporting personalized learning experiences and enhancing educational content delivery. By generating complex problem sets and providing tailored feedback, these models contribute to effective learning outcomes and student engagement [19]. This application underscores LLMs' potential to revolutionize educational methodologies and foster deeper understanding of complex subjects.

In autonomous systems, LLMs enhance task planning and decision-making in robotics and autonomous driving technologies [4]. This integration improves efficiency and safety, expanding applicability in real-world scenarios, from industrial automation to smart city infrastructure.

The impact of neural-symbolic systems and LLMs extends to creative industries, facilitating content generation and supporting creative processes [8]. By providing new tools for artistic expression, these systems enable creators to explore new frontiers in digital media and entertainment.

Cross-domain applications of neural-symbolic systems and LLMs highlight their transformative potential in enhancing productivity, innovation, and decision-making across fields. As these technologies advance, their real-world applications are anticipated to expand significantly, leading to transformative developments in sectors such as academia, law, and predictive analytics. This evolution is expected to enhance intelligent and adaptive systems, improving idea assessment in research, streamlining legal document generation, and integrating expert knowledge into predictive analytics for enhanced decision-making and risk assessment [2, 31, 34, 11, 21].

As illustrated in Figure 7, integrating language models in logical and mathematical problem-solving and cross-domain applications has demonstrated significant real-world impacts. The "Framework



(a) Framework for Trustworthiness Analysis of Language Models[68]

(b) Exploring Different Prompting Techniques in AI Tools[69]

(c) Response Variability in Standardized Evaluations[70]

Figure 7: Examples of Cross-Domain Applications and Real-World Impact

for Trustworthiness Analysis of Language Models" presents a systematic approach to evaluating the reliability of language model outputs across practical applications like sentiment analysis and truthfulness detection, emphasizing the importance of abstract model construction in ensuring dependability. Exploring "Different Prompting Techniques in AI Tools" provides insights into how various AI tools, such as ChatGPT and Spellburst, utilize distinct prompting strategies to enhance user interaction and task execution. Lastly, the "Response Variability in Standardized Evaluations" offers a comparative analysis of scores across different topics, underscoring the challenges in achieving genuine reasoning while balancing helpfulness with harmlessness. These examples collectively underscore the transformative potential of language models in solving complex problems and their cross-domain applicability, ultimately driving advancements in both theoretical and practical domains [68, 69, 70].

6 Challenges and Limitations

The deployment of neural networks, large language models (LLMs), and symbolic reasoning systems in AI is fraught with challenges that impact their performance, reliability, and integration. This section delves into the computational demands and efficiency required for AI systems, a critical factor in evaluating the feasibility of advanced AI solutions.

6.1 Computational Demands and Efficiency

AI systems integrating neural networks, LLMs, and symbolic reasoning face significant computational challenges. The high training and inference costs of large models limit their deployment in resource-constrained environments, with extensive parameter counts exacerbating these issues [3]. The generative nature of LLM outputs, which can produce a vast array of potential actions, complicates alignment with traditional strategies like Monte Carlo Tree Search (MCTS) [12]. In robotics, LLM-based methods struggle with computational efficiency, particularly in generating accurate motion functions for complex tasks [54]. The reliance on external graph reasoning APIs further increases computational demands due to variable quality and availability [23].

LLM outputs may blend factual information with hallucinations, complicating reliability and increasing computational demands [44]. Techniques like GFlowNet enhance diversity but do not significantly surpass reward-maximizing methods in accuracy [49]. Additionally, tracking long predicate relationship chains poses a challenge [46]. Benchmarks often fail to address the complexities of multi-robot coordination [56]. To optimize resource utilization and enhance model robustness, innovative strategies such as data augmentation and optimization algorithms are essential [3, 11, 2, 33].

6.2 Data Requirements and Quality

The effective integration of neural networks, LLMs, and symbolic reasoning systems hinges on stringent data requirements and quality assurance. Vulnerabilities such as fine-tuning and data poisoning highlight the need for secure datasets [6]. The static nature of knowledge in knowledge graphs (KGs) and reliance on LLMs for processing can lead to inadequacies in evaluation datasets, which may not capture dynamic real-world scenarios [71]. Noisy annotations in LLM-annotated datasets introduce inaccuracies, affecting performance, especially given the scarcity of annotated datasets [72].

In mathematical domains, synthetic data quality is crucial, necessitating validation to ensure accuracy [19]. Creativity evaluation subjectivity and initial prompt quality can lead to subpar outputs, emphasizing the need for reliable metrics [8]. Addressing data requirements and quality issues is vital for robust, effective AI technologies across domains, particularly in education, where academic integrity and responsible LLM use are paramount [2, 1, 35, 21].

6.3 Model Robustness and Generalization

Ensuring robustness and generalization in LLMs and neural-symbolic systems is crucial for their deployment across applications. These models are vulnerable to adversarial attacks and variability in outputs, leading to decision-making inconsistencies [73]. Generalization is constrained by training data limitations, affecting tasks requiring nuanced understanding [14]. The opacity of LLM decision-making complicates efforts to enhance robustness, as plausible yet incorrect outputs, or hallucinations, undermine reliability [23].

Improving robustness and generalization involves developing evaluation frameworks and incorporating structured reasoning techniques like knowledge graphs [12]. Feedback-driven learning mechanisms, such as Reinforcement Learning from Task Feedback (RLTF), can enhance adaptability in dynamic environments [36]. Addressing these challenges is essential for advancing LLMs and neural-symbolic systems, ensuring reliable performance across applications [2, 31, 11, 21].

6.4 Integration Challenges and Scalability

Integrating neural-symbolic systems with LLMs presents scalability challenges. A formal structure for describing and integrating these systems is essential for seamless operation [74]. The complexity of knowledge extraction and scalability of models further complicates integration [75]. Existing binary routing strategies for query handling do not fully leverage symbolic logic models (SLMs) and LLMs' strengths, particularly for complex queries [76]. Scalability issues also affect learning algorithms with high-dimensional input spaces, impacting efficiency and accuracy [77]. Ongoing research is needed to enhance scalability and integration capabilities, ensuring systems can tackle complex tasks across domains.

6.5 Ethical and Practical Implications

The deployment of LLMs and neural-symbolic systems raises ethical and practical challenges. Persuasive yet misleading outputs can impact informational integrity, necessitating robust ethical guidelines [57]. In education, nonsensical story elements produced by LLMs may confuse learners, highlighting the need for research to ensure AI systems enhance learning experiences [73]. Aligning LLMs with dynamic human values remains challenging, emphasizing the moral imperative for continual adaptation [24].

The proprietary nature of many models complicates transparency efforts, posing ethical challenges in high-stakes decision-making contexts [45]. Over-reliance on AI-generated documents may lead to biased judgments, underscoring the need for balanced AI integration [17]. LLM limitations, such as hallucination and accountability issues, pose ethical challenges distinct from traditional AI systems [65, 17, 34, 11, 47]. Addressing these limitations is crucial for accountability, minimizing biases, and enhancing transparency.

Practical challenges include designing user-centered interactions that enhance human-LLM collaboration and addressing diverse stakeholder needs. The availability of cost attributes for tools, essential for tool planning, highlights a practical limitation in AI deployment. Frameworks like Cost-Aware Tool Planning with Large Language Models (CATP-LLM) integrate cost considerations into planning, ensuring AI-generated plans are effective and viable [34, 78, 11].

Addressing the ethical and practical implications of LLMs and neural-symbolic systems is crucial for advancing AI responsibly. Prioritizing research, developing evaluation frameworks, and establishing ethical guidelines will facilitate transparency, accountability, and fairness, addressing challenges like bias, privacy, and information dissemination complexities. A human-centered perspective that considers diverse stakeholder needs can create a roadmap for integrating LLMs into society in beneficial and ethically sound ways [35, 2, 65].

7 Future Directions

Advancements in artificial intelligence, particularly through large language models (LLMs) and neural-symbolic systems, require strategic exploration of future directions to enhance these technologies. This section outlines key focus areas aimed at improving system capabilities and efficiency, ensuring effective real-world applications. The following subsection discusses strategies foundational for deploying LLMs across diverse contexts.

7.1 Enhancing Model Capabilities and Efficiency

Enhancing the capabilities and efficiency of LLMs and neural-symbolic systems is essential for their deployment in complex environments. Future research should focus on refining LLMs with domain-specific knowledge, addressing privacy concerns, and integrating them into clinical workflows to improve utility in healthcare [9]. Expanding frameworks to include complex social scenarios will improve model performance in dynamic contexts [4]. Optimizing question generation processes with open-source models and automated validation can make educational applications scalable [19].

Further integration with traditional NLP techniques across educational contexts could enhance interpretability and applicability, fostering effective learning experiences [14]. Developing practical dialogues between agents and humans through integrated supervised losses could improve human interpretability [7]. Recent studies emphasize the need for powerful, adaptable AI technologies that prioritize transparency and responsibility, integrating system-level improvements for better performance while ensuring ethical considerations are central [35, 79].

7.2 Integration with Domain-Specific Applications

Integrating LLMs and neural-symbolic systems with domain-specific applications presents significant opportunities for advancing AI capabilities. Future research should develop retrieval-augmented generation frameworks to enhance adaptability in dynamic environments [80]. This approach is valuable in finance and healthcare, where real-time data retrieval is critical. Adapting frameworks for diverse scientific disciplines can tailor AI systems to specific field needs, enhancing effectiveness [11]. Refining exposure scoring methodologies and developing regulatory frameworks are crucial for responsible AI deployment in employment contexts [81].

Exploring LLM applications in reasoning domains can assess broader AI applicability [82]. This integration enhances AI capabilities, enabling effective handling of complex tasks through improved reasoning processes. Addressing challenges in academic integrity and predictive analytics can lead to innovative solutions that adapt to user intent, enhance performance, and scale expert insights [41, 79, 1, 21, 47].

7.3 Advancements in Evaluation and Safety

Advancements in evaluating and ensuring the safety of LLMs and neural-symbolic systems are critical for reliable deployment. Multi-objective optimization techniques can enhance LLMs' ability to make informed decisions, improving performance while addressing privacy concerns [45]. Table 2 provides a detailed overview of the representative benchmarks employed in the evaluation of large language models and neural-symbolic systems, showcasing the diversity in domain applications, task formats, and performance metrics. Incorporating advanced evaluation frameworks is essential for assessing capabilities and limitations, providing systematic methodologies for benchmarking AI systems [83, 2, 11, 70, 35].

Integrating safety measures into LLM design and deployment is crucial for mitigating risks. This includes robust data privacy protocols and mechanisms to detect adversarial attacks, enhancing resilience in complex scenarios like emergency decision-making [6, 21]. These advancements facilitate the development of reliable applications, such as sophisticated plagiarism detection systems, enhancing academic integrity [1, 2].

AI-generated, for reference only.

Benchmark	Size	Domain	Task Format	Metric
BL[83]	1,000,000	Question Answering	Question Answering	Accuracy, F1-score
TDD[66]	1,000	Reading Comprehension	Multiple-Choice Question Answering	Success@k, Helpfulness
LLM-Factuality[84]	2,250	Text Summarization	Factuality Evaluation	Partial Correlation Coefficient, Factuality Errors
StAP-tutor[85]	148	Programming	Next-step Hint Generation	Clear, Helpful
MSE-LLM[16]	78	Mathematics	Question Answering	nDCG, P@10
EdTalk[47]	40	Education	Question Answering	BLEURT, Likert Scale
LLM-CPB[86]	104	Coding	Code Generation	Accuracy, Efficiency
T2T[87]	54,000	Table Reasoning	Question Answering	Accuracy, ROUGE-L

Table 2: This table presents a selection of representative benchmarks utilized in the evaluation of large language models (LLMs) and neural-symbolic systems. It includes various benchmarks across different domains, detailing the size, task format, and metrics used to assess performance. The table serves as a comprehensive overview of the current landscape in AI benchmarking, highlighting the diversity in evaluation methodologies.

7.4 Exploration of New Learning Paradigms

Exploring new learning paradigms is pivotal for advancing neural-symbolic systems and LLMs. Hybrid learning frameworks that integrate reinforcement learning with symbolic reasoning can enhance adaptability and efficiency in complex tasks [36]. Incorporating continual learning techniques maintains adaptability across evolving tasks, while inductive learning paradigms focus on refining logical theories [40, 46].

Novel data augmentation techniques, such as integrating diverse reasoning paths, can boost performance in logic-intensive domains like mathematical reasoning [38]. These paradigms are crucial for pushing the boundaries of AI technologies, facilitating advanced applications across domains like academic integrity and legal documentation [34, 11, 88, 2].

7.5 Neural Symbolic Systems in Educational Contexts

Integrating neural symbolic systems and LLMs into educational contexts holds promise for transforming teaching methodologies. These systems can revolutionize education by providing personalized learning experiences and facilitating deeper understanding through advanced reasoning capabilities [19]. Future applications could involve intelligent tutoring systems that adaptively respond to student inputs, enhancing education quality by fostering interactive learning environments [14, 66, 2, 50, 11].

Exploring collaborative learning experiences with neural symbolic systems presents opportunities for integrating machine learning and reasoning, addressing interpretability and accountability concerns [75, 11]. These systems can foster deeper understanding and reasoning among learners, contributing to explainable AI models that enhance educational interactions. Integrating these systems with educational technologies can revolutionize assessment tools, delivering nuanced evaluations of student performance and supporting academic integrity [14, 75, 1, 2]. By incorporating advanced reasoning capabilities, these tools provide insights into students' thought processes, offering educators a deeper understanding of student learning.

Exploring neural symbolic systems in educational contexts offers significant potential for transforming educational practices by integrating robust machine learning capabilities with principled knowledge representation, addressing interpretability and accountability issues [89, 75, 1, 11]. Harnessing AI power enables educators to create dynamic, personalized, and effective learning experiences, preparing students for modern challenges.

8 Conclusion

The convergence of neural networks, large language models (LLMs), and symbolic reasoning represents a pivotal advancement in artificial intelligence (AI), significantly enhancing problem-solving capabilities across diverse fields. This survey explores the synergy of these technologies, highlighting their combined potential to address intricate logical and mathematical challenges. Neural networks' adaptability pairs with the nuanced language comprehension and generation of LLMs, while symbolic reasoning offers the precision and structure necessary for sophisticated cognitive tasks.

The amalgamation of these technologies offers substantial advantages, including enhanced adaptability, interpretability, and efficiency, crucial for applications in healthcare, education, autonomous systems, and creative industries. Frameworks like GraphAgent-Reasoner and RLTF illustrate successful integrations of neural and symbolic paradigms, boosting AI systems' scalability and robustness. These frameworks facilitate the fusion of structured logic with dynamic learning environments, promoting the creation of more advanced and reliable AI applications.

Despite significant progress, challenges remain in optimizing computational efficiency, data quality, model robustness, and scalability. Overcoming these hurdles is essential for fully realizing the potential of neural-symbolic systems and ensuring their ethical application across various sectors. The exploration of new learning paradigms, such as hybrid frameworks and continual learning techniques, holds promise for advancing AI capabilities, enabling systems to execute complex reasoning tasks with increased precision and adaptability.

This conclusion underscores the importance of integrating neural networks, LLMs, and symbolic reasoning to boost AI's problem-solving prowess. Leveraging human feedback, methods like ALC3 can achieve oracle performance, refining modular AI systems and addressing the issues posed by noisy data. As the field progresses, formalizing neural-symbolic design patterns will offer a clearer framework for developing and comprehending these systems, paving the way for future innovations in AI technologies. This integration not only expands AI's problem-solving scope but also fosters the development of more interpretable, reliable, and efficient AI systems across various applications.

References

[1] Suriya Prakash Jambunathan, Ashwath Shankarnarayan, and Parijat Dube. Convnlp: Image-based ai text detection, 2024.

[2] Mujahid Ali Quidwai, Chunhui Li, and Parijat Dube. Beyond black box ai-generated plagiarism detection: From sentence to document level, 2023.

[3] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[4] Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. Richelieu: Self-evolving llm-based agents for ai diplomacy, 2024.

[5] Tuan Bui, Oanh Tran, Phuong Nguyen, Bao Ho, Long Nguyen, Thang Bui, and Tho Quan. Cross-data knowledge graph construction for llm-enabled educational question-answering system: A case study at hcmut, 2024.

[6] Minze Chen, Zhenxiang Tao, Weitong Tang, Tingxin Qin, Rui Yang, and Chunli Zhu. Enhancing emergency decision-making with knowledge graphs and large language models, 2023.

[7] Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. *Advances in neural information processing systems*, 30, 2017.

[8] Jeremy Straub and Zach Johnson. Initial development and evaluation of the creative artificial intelligence through recurring developments and determinations (cairdd) system, 2024.

[9] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, and Bing Shi. Chatgpt for shaping the future of dentistry: The potential of multi-modal large language model, 2023.

[10] Chandan Singh, Jeevana Priya Inala, Michel Galley, Rich Caruana, and Jianfeng Gao. Rethinking interpretability in the era of large language models, 2024.

[11] Yi Xu, Bo Xue, Shuqian Sheng, Cheng Deng, Jiaxin Ding, Zanwei Shen, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. Good idea or not, representation of llm could tell, 2024.

[12] Di Zhang, Xiaoshui Huang, Dongzhan Zhou, Yuqiang Li, and Wanli Ouyang. Accessing gpt-4 level mathematical olympiad solutions via monte carlo tree self-refine with llama-3 8b, 2024.

[13] Zixian Huang, Wenhao Zhu, Gong Cheng, Lei Li, and Fei Yuan. Mindmerger: Efficient boosting llm reasoning in non-english languages, 2024.

[14] Lishan Zhang, Han Wu, Xiaoshan Huang, Tengfei Duan, and Hanxiang Du. Automatic deductive coding in discourse analysis: an application of large language models in learning analytics, 2024.

[15] Chao Zhang, Xuechen Liu, Katherine Ziska, Soobin Jeon, Chi-Lin Yu, and Ying Xu. Mathemyths: Leveraging large language models to teach mathematical language through child-ai co-creative storytelling, 2024.

[16] Ankit Satpute, Noah Giessing, Andre Greiner-Petter, Moritz Schubotz, Olaf Teschke, Akiko Aizawa, and Bela Gipp. Can llms master math? investigating large language models on math stack exchange, 2024.

[17] Jingyu Liu, Jiaen Lin, and Yong Liu. How much can rag help the reasoning of llm?, 2024.

[18] Denis Havlik and Marcelo Pias. Common errors in generative ai systems used for knowledge extraction in the climate action domain, 2024.

[19] Vedant Shah, Dingli Yu, Kaifeng Lyu, Simon Park, Jiatong Yu, Yinghui He, Nan Rosemary Ke, Michael Mozer, Yoshua Bengio, Sanjeev Arora, and Anirudh Goyal. Ai-assisted generation of difficult math questions, 2025.

[20] Yuzi Yan, Jialian Li, Yipin Zhang, and Dong Yan. Exploring the llm journey from cognition to expression with linear representations, 2024.

[21] Phoebe Jing, Yijing Gao, Yuanhang Zhang, and Xianlong Zeng. Translating expert intuition into quantifiable features: Encode investigator domain knowledge via llm for enhanced predictive analytics, 2024.

[22] Jiawei Zhang. Graph-toolformer: To empower llms with graph reasoning ability via prompt augmented by chatgpt, 2023.

[23] Linhao Luo, Zicheng Zhao, Chen Gong, Gholamreza Haffari, and Shirui Pan. Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models, 2024.

[24] Yuwei Hu, Runlin Lei, Xinyi Huang, Zhewei Wei, and Yongchao Liu. Scalable and accurate graph reasoning with llm-based multi-agents, 2024.

[25] Ante Wang, Linfeng Song, Ye Tian, Baolin Peng, Lifeng Jin, Haitao Mi, Jinsong Su, and Dong Yu. Self-consistency boosts calibration for math reasoning, 2024.

[26] Junlin Wang, Siddhartha Jain, Dejiao Zhang, Baishakhi Ray, Varun Kumar, and Ben Athiwaratkun. Reasoning in token economies: Budget-aware evaluation of llm reasoning strategies, 2024.

[27] Michael Xieyang Liu, Frederick Liu, Alexander J. Fiannaca, Terry Koo, Lucas Dixon, Michael Terry, and Carrie J. Cai. "we need structured output": Towards user-centered constraints on large language model output, 2024.

[28] Subir Majumder, Lin Dong, Fatemeh Doudi, Yuting Cai, Chao Tian, Dileep Kalathi, Kevin Ding, Anupam A. Thatte, Na Li, and Le Xie. Exploring the capabilities and limitations of large language models in the electric energy sector, 2024.

[29] Andrew M. Bean, Karolina Korgul, Felix Krones, Robert McCraith, and Adam Mahdi. Do large language models have shared weaknesses in medical question answering?, 2024.

[30] Ming Li, Yanhong Li, and Tianyi Zhou. What happened in llms layers when trained for fast vs. slow thinking: A gradient perspective, 2024.

[31] Ji Qi, Kaixuan Ji, Xiaozhi Wang, Jifan Yu, Kaisheng Zeng, Lei Hou, Juanzi Li, and Bin Xu. Mastering the task of open information extraction with large language models and consistent reasoning environment, 2023.

[32] Milena Chadimová, Eduard Jurášek, and Tomáš Kliegr. Meaningless is better: hashing bias-inducing words in llm prompts improves performance in logical reasoning and statistical learning, 2024.

[33] Bosheng Ding, Chengwei Qin, Ruochen Zhao, Tianze Luo, Xinze Li, Guizhen Chen, Wenhan Xia, Junjie Hu, Anh Tuan Luu, and Shafiq Joty. Data augmentation using large language models: Data perspectives, learning paradigms and challenges, 2024.

[34] Jakub Harasta, Tereza Novotná, and Jaromir Savelka. It cannot be right if it was written by ai: On lawyers' preferences of documents perceived as authored by an llm vs a human, 2024.

[35] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[36] Yingqiang Ge, Wenyue Hua, Kai Mei, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang, et al. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems*, 36:5539–5568, 2023.

[37] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling, 2024.

[38] Zheng Yuan, Hongyi Yuan, Chengpeng Li, Guanting Dong, Keming Lu, Chuanqi Tan, Chang Zhou, and Jingren Zhou. Scaling relationship on learning mathematical reasoning with large language models, 2023.

[39] Matthieu Lin, Jenny Sheng, Andrew Zhao, Shenzhi Wang, Yang Yue, Yiran Wu, Huan Liu, Jun Liu, Gao Huang, and Yong-Jin Liu. Llm-based optimization of compound ai systems: A survey, 2024.

[40] Junhao Zheng, Shengjie Qiu, Chengming Shi, and Qianli Ma. Towards lifelong learning of large language models: A survey, 2024.

[41] Zhuocheng Shen. Llm with tools: A survey. *arXiv preprint arXiv:2409.18807*, 2024.

[42] Han Meng, Yitian Yang, Yunan Li, Jungup Lee, and Yi-Chieh Lee. Exploring the potential of human-llm synergy in advancing qualitative analysis: A case study on mental-illness stigma, 2024.

[43] Zhihu Wang, Shiwan Zhao, Yu Wang, Heyuan Huang, Sitao Xie, Yubo Zhang, Jiaxin Shi, Zhixing Wang, Hongyan Li, and Junchi Yan. Re-task: Revisiting llm tasks from capability, skill, and knowledge perspectives, 2024.

[44] Mohammed Saeed, Nicola De Cao, and Paolo Papotti. Querying large language models with sql, 2023.

[45] Kai Zhang, Liqian Peng, Congchao Wang, Alec Go, and Xiaozhong Liu. Llm cascade with multi-objective optimal consideration, 2024.

[46] João Pedro Gandarela, Danilo S. Carvalho, and André Freitas. Inductive learning of logical theories with llms: An expressivity-graded analysis, 2025.

[47] Bhashithe Abeysinghe and Ruhan Circi. The challenges of evaluating llm applications: An analysis of automated, human, and llm-based approaches, 2024.

[48] Eva Eigner and Thorsten Händler. Determinants of llm-assisted decision-making, 2024.

[49] Ryoichi Takase, Masaya Tsunokake, Yuta Tsuchiya, and Shota Inuzuka. Gflownet fine-tuning for diverse correct solutions in mathematical reasoning tasks, 2024.

[50] Lars Krupp, Jonas Bley, Isacco Gobbi, Alexander Geng, Sabine Müller, Sungho Suh, Ali Moghiseh, Arcesio Castaneda Medina, Valeria Bartsch, Artur Widera, Herwig Ott, Paul Lukowicz, Jakob Karolus, and Maximilian Kiefer-Emmanouilidis. Llm-generated tips rival expert-created tips in helping students answer quantum-computing questions, 2024.

[51] Yeonhong Park, Jake Hyun, SangLyul Cho, Bonggeun Sim, and Jae W. Lee. Any-precision llm: Low-cost deployment of multiple, different-sized llms, 2024.

[52] Tianyang Xu, Shujin Wu, Shizhe Diao, Xiaoze Liu, Xingyao Wang, Yangyi Chen, and Jing Gao. Sayself: Teaching llms to express confidence with self-reflective rationales, 2024.

[53] Haokun Liu, Yaonan Zhu, Kenji Kato, Izumi Kondo, Tadayoshi Aoyama, and Yasuhisa Hasegawa. Llm-based human-robot collaboration framework for manipulation tasks, 2023.

[54] Vineet Bhat, Ali Umut Kaypak, Prashanth Krishnamurthy, Ramesh Karri, and Farshad Khorrami. Grounding llms for robot task planning using closed-loop state feedback, 2024.

[55] Hamed Babaei Giglou, Jennifer D'Souza, and Sören Auer. Llms4ol: Large language models for ontology learning, 2023.

[56] Yongchao Chen, Jacob Arkin, Yang Zhang, Nicholas Roy, and Chuchu Fan. Scalable multi-robot collaboration with large language models: Centralized or decentralized systems?, 2024.

[57] Vahid Ashrafimoghari, Necdet Gürkan, and Jordan W. Suchow. Evaluating large language models on the gmat: Implications for the future of business education, 2024.

[58] Prateek Verma. Wavelet gpt: Wavelet inspired large language models, 2025.

AI-generated, for reference only.

[59] John J. Horton. Large language models as simulated economic agents: What can we learn from homo silicus?, 2023.

[60] Zhiheng Xi, Dingwen Yang, Jixuan Huang, Jiafu Tang, Guanyu Li, Yiwen Ding, Wei He, Boyang Hong, Shihan Do, Wenyu Zhan, Xiao Wang, Rui Zheng, Tao Ji, Xiaowei Shi, Yitao Zhai, Rongxiang Weng, Jingang Wang, Xunliang Cai, Tao Gui, Zuxuan Wu, Qi Zhang, Xipeng Qiu, Xuanjing Huang, and Yu-Gang Jiang. Enhancing llm reasoning via critique models with test-time and training-time supervision, 2024.

[61] Qixiang Fang, Daniel L. Oberski, and Dong Nguyen. Patch! psychometrics-assisted benchmarking of large language models: A case study of proficiency in 8th grade mathematics, 2024.

[62] Haoxiong Liu, Yifan Zhang, Yifan Luo, and Andrew Chi-Chih Yao. Augmenting math word problems via iterative question composing, 2024.

[63] Zhenyu Wu, Meng Jiang, and Chao Shen. Get an a in math: Progressive rectification prompting, 2023.

[64] Danqing Wang, Zhuorui Ye, Fei Fang, and Lei Li. Cooperative strategic planning enhances reasoning capabilities in large language models, 2024.

[65] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. Navigating llm ethics: Advancements, challenges, and future directions, 2024.

[66] Menna Fateen and Tsunenori Mine. Developing a tutoring dialog dataset to optimize llms for educational use, 2024.

[67] James Prather, Paul Denny, Juho Leinonen, David H. Smith IV au2, Brent N. Reeves, Stephen MacNeil, Brett A. Becker, Andrew Luxton-Reilly, Thezyrie Amarouche, and Bailey Kimmel. Interactions with prompt problems: A new way to teach programming with large language models, 2024.

[68] Da Song, Xuan Xie, Jiayang Song, Derui Zhu, Yuheng Huang, Felix Juefei-Xu, and Lei Ma. Luna: A model-based universal analysis framework for large language models, 2024.

[69] Hariharan Subramonyam, Roy Pea, Christopher Lawrence Pondoc, Maneesh Agrawala, and Colleen Seifert. Bridging the gulf of envisioning: Cognitive design challenges in llm interfaces, 2024.

[70] Timothy R. McIntosh, Teo Susnjak, Nalin Arachchilage, Tong Liu, Paul Watters, and Malka N. Halgamuge. Inadequacies of large language model benchmarks in the era of generative artificial intelligence, 2024.

[71] Ernests Lavrinovics, Russa Biswas, Johannes Bjerva, and Katja Hose. Knowledge graphs, large language models, and hallucinations: An nlp perspective, 2024.

[72] Kristina Gligorić, Tijana Zrnic, Cinoo Lee, Emmanuel J. Candès, and Dan Jurafsky. Can unconfident llm annotations be used for confident conclusions?, 2025.

[73] Yuyou Gan, Yong Yang, Zhe Ma, Ping He, Rui Zeng, Yiming Wang, Qingming Li, Chunyi Zhou, Songze Li, Ting Wang, Yunjun Gao, Yingcai Wu, and Shouling Ji. Navigating the risks: A survey of security, privacy, and ethics threats in llm-based agents, 2024.

[74] Till Mossakowski. Modular design patterns for neural-symbolic integration: refinement and combination, 2022.

[75] Artur d'Avila Garcez, Marco Gori, Luis C Lamb, Luciano Serafini, Michael Spranger, and Son N Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv preprint arXiv:1905.06088*, 2019.

[76] Adarsh MS, Jithin VG, and Ditto PS. Efficient hybrid inference for llms: Reward-based token modelling with selective cloud assistance, 2024.

[77] Alaia Solko-Breslin, Seewon Choi, Ziyang Li, Neelay Velingker, Rajeev Alur, Mayur Naik, and Eric Wong. Data-efficient learning with neural programs, 2024.

[78] Duo Wu, Jinghe Wang, Yuan Meng, Yanning Zhang, Le Sun, and Zhi Wang. Catp-llm: Empowering large language models for cost-aware tool planning, 2024.

[79] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. Llm inference serving: Survey of recent advances and opportunities, 2024.

[80] Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhattacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, et al. From generation to judgment: Opportunities and challenges of llm-as-a-judge. *arXiv preprint arXiv:2411.16594*, 2024.

[81] Qin Chen, Jinfeng Ge, Huaqing Xie, Xingcheng Xu, and Yanqing Yang. Large language models at work in china's labor market, 2023.

[82] Shiyang Li, Jianshu Chen, Yelong Shen, Zhiyu Chen, Xinlu Zhang, Zekun Li, Hong Wang, Jing Qian, Baolin Peng, Yi Mao, Wenhu Chen, and Xifeng Yan. Explanations from large language models make small reasoners better, 2022.

[83] Kun Zhou, Yutao Zhu, Zhipeng Chen, Wentong Chen, Wayne Xin Zhao, Xu Chen, Yankai Lin, Ji-Rong Wen, and Jiawei Han. Don't make your llm an evaluation benchmark cheater. *arXiv preprint arXiv:2311.01964*, 2023.

[84] Xue-Yong Fu, Md Tahmid Rahman Laskar, Cheng Chen, and Shashi Bhushan TN. Are large language models reliable judges? a study on the factuality evaluation capabilities of llms, 2023.

[85] Lianne Roest, Hieke Keuning, and Johan Jeuring. Next-step hint generation for introductory programming using large language models, 2023.

[86] Lincoln Murr, Morgan Grainger, and David Gao. Testing llms on code generation with varying levels of prompt specificity, 2023.

[87] Naihao Deng, Zhenjie Sun, Ruiqi He, Aman Sikka, Yulong Chen, Lin Ma, Yue Zhang, and Rada Mihalcea. Tables as texts or images: Evaluating the table reasoning ability of llms and mllms, 2024.

[88] Zach Johnson and Jeremy Straub. Development of regai: Rubric enabled generative artificial intelligence, 2024.

[89] Sakshi Mahendru and Tejul Pandit. Venn diagram prompting : Accelerating comprehension with scaffolding effect, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.