Task Specification and Planning in Robotics: A Survey

www.surveyx.cn

Abstract

This survey paper explores the interdisciplinary advancements in task specification, planning, and execution within robotics, emphasizing the integration of robotics, artificial intelligence (AI), and natural language processing (NLP). The paper highlights the transformative role of these interdisciplinary approaches in enhancing the adaptability and efficiency of robotic systems. Key frameworks such as RoboGPT and PsALM exemplify advancements in intelligent agent capabilities and mission design precision, respectively. The integration of Graph Neural Networks (GNNs) and innovative frameworks like RAHL demonstrate improved decision-making and task planning through advanced computational models. The survey underscores the importance of collaborative frameworks in multi-human single-robot interactions, as evidenced by empirical results showcasing enhanced team performance. Additionally, the application of Large Language Models (LLMs) in tool-object manipulation and manoeuvrability-driven approaches represents significant progress in robotic capabilities. The survey concludes by emphasizing the critical role of interdisciplinary strategies in fostering intelligent, adaptable, and efficient robotic systems capable of operating in complex and dynamic environments, thereby facilitating their deployment across diverse applications.

1 Introduction

1.1 Interdisciplinary Nature of Task Specification and Planning

Task specification and planning in robotics necessitate a multidisciplinary approach, integrating robotics, artificial intelligence (AI), natural language processing (NLP), and knowledge representation to create autonomous systems capable of performing complex tasks in varied environments [1]. This integration addresses challenges in robotic task planning, particularly the need for adaptability to novel problems, as seen in robotic cooking scenarios where limited knowledge of actions and environments can impede effective task execution [2].

The development of household embodied agents capable of selecting substitute objects for various tasks emphasizes the need for combining robotics, AI, and commonsense reasoning to enhance decision-making and planning capabilities [3]. Additionally, incorporating language-gesture interactions in robotic systems reveals the limitations of traditional verbal instruction methods, advocating for a holistic approach that utilizes multiple modalities to improve task communication and planning [4].

Advancements in vision-language models (VLMs) exemplify this interdisciplinary confluence, integrating classical planning techniques to enhance robotic task planning [5]. Large Language Models (LLMs) are pivotal in this context, providing advanced reasoning capabilities essential for effective human-robot interaction and autonomous task planning [6]. The synergy of NLP techniques within Business Process Management (BPM) further illustrates this interdisciplinary integration, allowing natural language commands to be translated into formal task specifications, such as linear temporal logic (LTL), enabling robots to interpret and execute human instructions accurately.

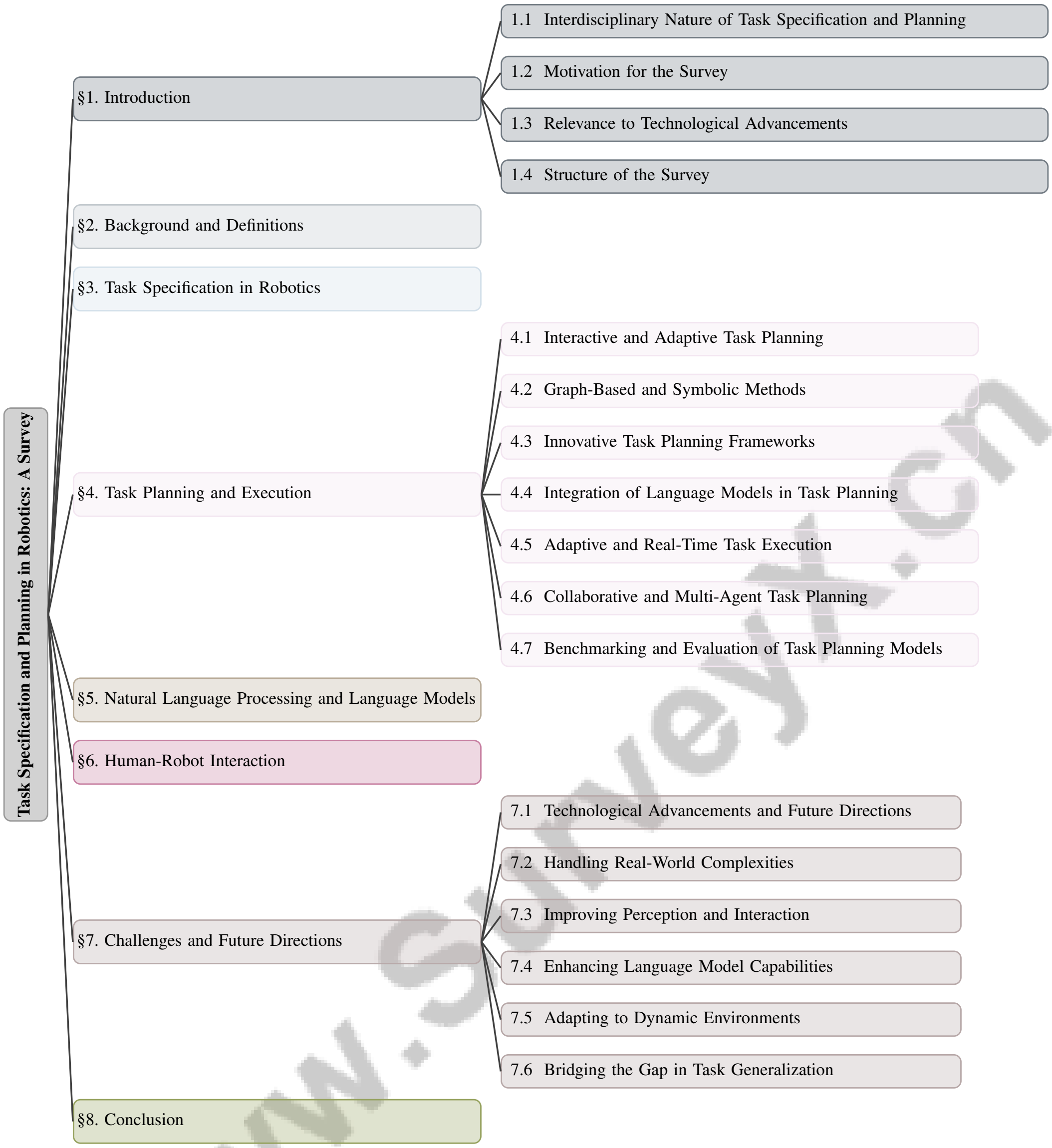

Task Specification and Planning in Robotics: A Survey
§1. Introduction
1.1 Interdisciplinary Nature of Task Specification and Planning
1.2 Motivation for the Survey
1.3 Relevance to Technological Advancements
1.4 Structure of the Survey
§2. Background and Definitions
§3. Task Specification in Robotics
§4. Task Planning and Execution
4.1 Interactive and Adaptive Task Planning
4.2 Graph-Based and Symbolic Methods
4.3 Innovative Task Planning Frameworks
4.4 Integration of Language Models in Task Planning
4.5 Adaptive and Real-Time Task Execution
4.6 Collaborative and Multi-Agent Task Planning
4.7 Benchmarking and Evaluation of Task Planning Models
§5. Natural Language Processing and Language Models
§6. Human-Robot Interaction
§7. Challenges and Future Directions
7.1 Technological Advancements and Future Directions
7.2 Handling Real-World Complexities
7.3 Improving Perception and Interaction
7.4 Enhancing Language Model Capabilities
7.5 Adapting to Dynamic Environments
7.6 Bridging the Gap in Task Generalization
§8. Conclusion

Figure 1: chapter structure

In industrial applications, the demand for flexible, skill-based robot control systems underscores the necessity for an interdisciplinary approach that merges knowledge representation, task planning, and execution strategies [7]. The pursuit of artificial general intelligence (AGI) is accelerated by Multimodal Large Language Models (MLLMs), emphasizing the need for human-level planning capabilities crucial for informed decision-making in complex environments [8]. This interdisciplinary integration is vital for fostering innovations that enhance the efficiency and adaptability of robotic systems, ensuring that human-robot collaboration aligns with human performance metrics [9].

1.2 Motivation for the Survey

This survey addresses critical gaps in the application of Large Language Models (LLMs) within robotic environments, particularly their limitations in generating executable task plans due to insufficient grounding in physical contexts, which often leads to impractical action sequences [10]. Current methodologies struggle to adapt to unforeseen situations in open-world environments, underscoring the need for more robust and adaptable solutions [5]. Existing robotic task planning methods also face inefficiencies and inaccuracies in dynamic or unstructured environments, necessitating the development of flexible frameworks [11].

The survey is motivated by the pressing need to enhance data privacy in machine learning models, particularly when handling sensitive information within robotic systems [12]. Additionally, there is a notable gap in the versatility of models within Business Process Management (BPM) that can effectively comprehend process-related texts, highlighting a deficiency in the field [13]. The challenge of developing task planning methods that respect privacy, security, and access control while enabling real-time learning and adaptation in robots further underscores the necessity of this survey [14].

Moreover, there is a significant need for high-level autonomy in robots capable of adapting to unpredictable environments, as previous methods have struggled with vague task specifications [15]. The creation of robust task planning systems that allow robots to learn and adapt to new environments without extensive prior knowledge is urgent [16]. The inefficiency and complexity of existing task planning methods for multi-drone systems highlight the necessity for more adaptable solutions [17].

Finally, the survey seeks to address limitations in robotic systems' self-correction abilities during task execution, which restricts their adaptability in dynamic environments [18]. By exploring these challenges, the survey aims to contribute to the advancement of effective planning and reasoning capabilities in robotics, enhancing human-robot interaction across various application domains. The integration of low-level feasibility checks into high-level task planning is crucial for improving robotic performance in dynamic environments [19], while the reuse of task plans across varied environments is essential for industries reliant on automation [1]. Addressing the challenge of effectively communicating tasks to robots using simple human instructions is vital, as existing methods struggle with ambiguous verbal instructions [4]. The survey emphasizes the importance of realistic task scenarios and complex decision-making in advancing human-level planning capabilities in MLLMs [8], as well as enhancing robots' abilities to generate task plans from knowledge graphs like the Functional Object-Oriented Network (FOON) to improve adaptability in robotic cooking [2].

1.3 Relevance to Technological Advancements

The survey's relevance to technological advancements is underscored by the integration of LLMs and VLMs into task planning systems within robotics, significantly enhancing robots' abilities to process natural language and execute complex instructions. These advancements improve decision-making and adaptability in dynamic environments, aligning with the evolving landscape of robotics and AI. Frameworks that combine traditional symbolic planning with neural networks and LLMs, as described by Hemken et al., exemplify the creation of flexible, adaptive planning systems capable of learning from context, further underscoring the survey's relevance to current technological progress [14].

The emergence of VLMs as a promising alternative to traditional methods reflects recent technological advancements in robotics and AI, enabling more sophisticated interaction and planning capabilities [11]. Additionally, advancements in drone technology necessitate improved human-machine interaction through innovative task planning methods, emphasizing the survey's relevance to these developments [17].

Technological progress is further exemplified by methods that enable robots to learn new skills from demonstration videos, a critical aspect of the evolving landscape of robotics and AI [15]. The HiCRISP framework addresses existing limitations by allowing robots to correct errors within individual steps during task execution, showcasing its relevance to current advancements [18].

Furthermore, the ability to generate task trees for recipes not explicitly present in the FOON enhances the flexibility and applicability of robotic cooking systems, demonstrating the survey's alignment with technological advancements [2]. The integration of discrete high-level reasoning and continuous low-level reasoning to improve hybrid planning in robotics addresses the complexity and variability of robotic tasks, emphasizing the need for new methods to effectively handle these challenges [19]. These developments collectively illustrate the survey's alignment with the rapidly advancing landscape of robotics and AI, offering insights into the integration of cutting-edge technologies for improved task planning and execution.

1.4 Structure of the Survey

This survey paper is organized into several key sections to systematically explore the interdisciplinary fields and technologies involved in task specification, task planning, natural language processing, robotics, human-robot interaction, and language models. The paper begins with an introduction that outlines the scope and significance of the survey, emphasizing its interdisciplinary nature and relevance to current technological advancements. Following this, the background and definitions section provides an overview of the core concepts and terminologies that underpin the study, setting a foundation for understanding the subsequent discussions.

The survey is then divided into thematic sections, starting with a detailed examination of task specification in robotics, discussing methodologies and technologies used, including natural language-based task specification and the integration of multimodal approaches. This is followed by an exploration of task planning and execution, highlighting strategies and algorithms utilized in robotics, focusing on interactive and adaptive task planning, graph-based and symbolic methods, and innovative frameworks. The integration of language models in task planning and the importance of adaptive and real-time task execution are also addressed.

Next, the paper delves into advancements in natural language processing and language models, exploring their role in facilitating human-robot interaction and task planning. This section discusses the integration of language models with knowledge-based systems, advancements in vision-language models, and the challenges associated with generating factual and consistent responses.

The dynamics of human-robot interaction are then analyzed, focusing on communication, collaboration, and task-sharing strategies that enhance collaborative task execution. The survey concludes with a comprehensive discussion of challenges and future directions in the field of robotics, highlighting significant technological advancements and potential research avenues that could effectively address current limitations. It emphasizes the need for improved robot perception, interaction, and adaptability in dynamic environments, particularly through enhanced task specification frameworks that reduce user burden, adaptive task planning that aligns with human preferences, and innovative relevance frameworks that streamline human-robot collaboration. Additionally, the survey underscores the importance of developing robust methods for embodied vision-language planning and exploring novel interaction modalities, such as brain-computer interfaces, to facilitate seamless operation in real-world scenarios [20, 21, 22, 12, 23].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts of Task Specification and Planning

Task specification and planning in robotics involve translating high-level objectives into executable actions within complex environments, integrating NLP, visual perception, and action execution to develop Vision-Language-Action (VLA) models [24]. A significant challenge is bridging high-level reasoning with practical execution, requiring robust frameworks [25]. LLMs often struggle with long-horizon tasks due to limited planning competence [26]. Employing LLMs in task planning entails interpreting user commands via prompt-driven methods, essential for converting natural language instructions into structured representations [17]. Techniques like Linear Temporal Logic (LTL) capture temporal and logical dependencies, enabling systematic execution [27]. However, reliance on fixed templates in LLM-based Task and Motion Planning (TAMP) can lead to semantically flawed plans, highlighting the need for more dynamic models [28]. In large-scale domains with partial observations, such as household environments, managing objects and their spatial relations presents computational challenges [29]. The computational intractability of existing planning methods, due to expansive state spaces, complicates feasible plan generation [30].

Integrating symbolic planning with neural networks allows robots to adapt task execution based on dynamic constraints, exemplifying a core concept in task planning [14]. Generating machine-readable problem descriptions from linguistic instructions and scene observations enables symbolic planners to derive valid plans [31]. Iterative self-refinement of LLMs to generate feasible action plans for complex tasks further illustrates key task planning concepts [32]. Specification mining algorithms that generate parametric GSTL formulas from demonstration videos facilitate automated task planning, showcasing the integration of visual data into task planning frameworks [15]. Traditional rule-driven and learning-based methods' limitations, lacking adaptability in dynamic environments, highlight the necessity for innovative task planning approaches [11].

Incorporating human preferences into robotic task planning is vital for optimizing collaboration outcomes, emphasizing the human-centric aspect of task specification [9]. Balancing model accuracy

AI-generated, for reference only.

with data privacy underscores the significance of privacy in robotic task planning [12]. Furthermore, the need for versatile instruments capable of handling various tasks using LLMs in Business Process Management (BPM) emphasizes the flexibility required in task planning frameworks [13].

2.2 Natural Language Processing and Language Models

NLP and language models are crucial for enhancing human-robot interaction by enabling robots to comprehend and generate human language effectively. NLP encompasses computational techniques for analyzing and generating human language, facilitating systems that process natural language inputs [33]. Advancements in machine learning, particularly pre-training and fine-tuning, have significantly influenced NLP, treating tasks as text generation problems [6]. LLMs, as essential tools in advancing NLP, predict subsequent words in a sequence, generating coherent and contextually relevant text. Their integration into robotic systems enhances autonomous behavior planning and execution under textual instructions, employing multimodal sensory feedback for failure detection. This capability is crucial for robots executing high-level instructions grounded in reality, as demonstrated by platforms like SkiROS2, which utilize a world model and task manager for planning and execution [7].

NLP and language models play a critical role in understanding and generating language, vital for effective task planning and execution in robotics. Key challenges include modeling vision and language, ensuring task generalization, and developing robust frameworks for real-world deployment in embodied vision-language planning tasks [34]. Integrating multimodal datasets comprising text and images offers a comprehensive approach to enhancing language models' capabilities in processing diverse inputs. Effectively interpreting human guidance poses a significant challenge, as it is often less direct than specified rewards or demonstrations. This highlights the importance of developing frameworks that categorize LLM capabilities, focusing on strengths in NLP and data analysis to minimize misunderstandings in user interactions [6]. Novel frameworks for categorizing NLP applications across various fields, such as management, further illustrate the versatility and applicability of these technologies [33].

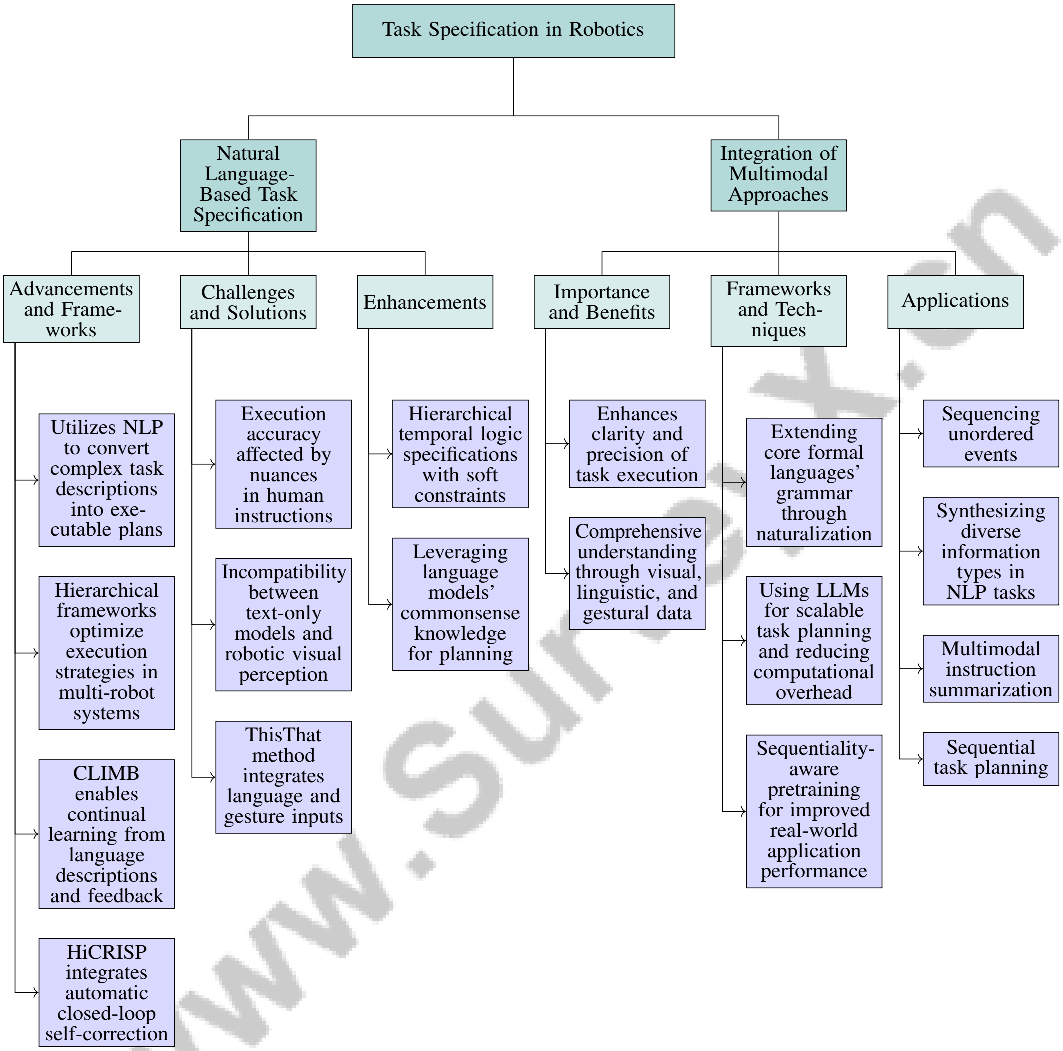
2.3 Human-Robot Interaction Dynamics

Human-robot interaction (HRI) is fundamental to task execution by robotic systems, emphasizing seamless communication and collaboration between humans and robots. HRI principles focus on understanding human intentions, interpreting natural language commands, and adapting to human preferences, crucial for optimizing task execution in human-centric environments. Integrating advanced NLP techniques, particularly large pre-trained language models (LLMs) like BERT and multimodal models such as GPT-4V, significantly enhances robots' ability to interpret and respond to human instructions. This advancement fosters more intuitive interactions and improves robots' task planning capabilities in complex settings by merging natural language comprehension with visual perception [35, 36, 37].

The importance of HRI in task execution is underscored by the need for robots to adapt to dynamic environments, collaborating with humans to accomplish complex tasks. This adaptability requires robots to understand human social cues and adjust their actions based on real-time feedback from collaborators [9]. Developing frameworks that incorporate multimodal interactions, merging verbal and non-verbal communication, is essential for enhancing task instruction clarity and improving overall human-robot collaboration efficiency [4]. Additionally, HRI extends to designing robotic systems that learn from human demonstrations and adapt their behavior accordingly. This capability is vital for aligning robots' actions with human expectations and preferences, fostering harmonious interactions [15]. Incorporating human feedback into robotic task planning and execution processes enhances adaptability and ensures alignment with human goals and objectives.

3 Task Specification in Robotics

Addressing the complexities of task specification in robotics requires innovative methodologies, notably the use of natural language as a bridge between human intent and robotic action. This approach facilitates intuitive interactions and enhances robotic systems' adaptability by enabling them to interpret and execute tasks based on human-like instructions. As illustrated in Figure 2, the hierarchical structure of task specification in robotics is depicted, emphasizing the roles of natural language-based approaches and multimodal integration. This figure categorizes advancements, challenges, and enhancements in natural language-based task specification, while also outlining the importance, frameworks, techniques, and applications of integrating multimodal approaches to improve robotic task execution. The following subsection explores these advancements, frameworks, and challenges in natural language-based task specification.



Task Specification in Robotics
Natural Language-Based Task Specification
Integration of Multimodal Approaches
Advancements and Frameworks
Challenges and Solutions
Enhancements
Importance and Benefits
Frameworks and Techniques
Applications
Utilizes NLP to convert complex task descriptions into executable plans
Hierarchical frameworks optimize execution strategies in multi-robot systems
CLIMB enables continual learning from language descriptions and feedback
HiCRISP integrates automatic closed-loop self-correction
Execution accuracy affected by nuances in human instructions
Incompatibility between text-only models and robotic visual perception
ThisThat method integrates language and gesture inputs
Hierarchical temporal logic specifications with soft constraints
Leveraging language models' commonsense knowledge for planning
Enhances clarity and precision of task execution
Comprehensive understanding through visual, linguistic, and gestural data
Extending core formal languages' grammar through naturalization
Using LLMs for scalable task planning and reducing computational overhead
Sequentiality-aware pretraining for improved real-world application performance
Sequencing unordered events
Synthesizing diverse information types in NLP tasks
Multimodal instruction summarization
Sequential task planning

Figure 2: This figure illustrates the hierarchical structure of task specification in robotics, highlighting the roles of natural language-based approaches and multimodal integration. The figure categorizes advancements, challenges, and enhancements in natural language-based task specification and outlines the importance, frameworks, techniques, and applications of integrating multimodal approaches to improve robotic task execution.

3.1 Natural Language-Based Task Specification

Natural language-based task specification represents a significant advancement in human-robot interaction, allowing robots to interpret and execute tasks from human-like instructions. This approach utilizes natural language processing (NLP) to convert complex task descriptions into executable plans, enhancing robotic systems' adaptability and responsiveness [17]. In multi-robot systems, natural language facilitates task specification, addressing reliability and performance challenges through hierarchical frameworks that optimize execution strategies and minimize collaboration wait times [38].

AI-generated, for reference only.

Frameworks such as CLIMB demonstrate the advantages of natural language by enabling robots to build and refine domain models through continual learning from language descriptions and execution feedback [16]. The HiCRISP framework further illustrates the potential by integrating automatic closed-loop self-correction for real-time error correction during task execution [18].

Challenges persist, particularly regarding execution accuracy, as language models may struggle with human instruction nuances. The incompatibility between text-only models and robotic visual perception complicates the process [34]. The This&That method addresses this by integrating language and gesture inputs, improving video generation for robot planning [4].

Natural language interfaces are further enhanced by hierarchical temporal logic specifications incorporating soft constraints, allowing agents to dynamically select sub-tasks that meet global objectives [39]. Additionally, leveraging language models' commonsense knowledge to identify relevant objects simplifies planning, showcasing natural language's potential in streamlining task specification [2].

3.2 Integration of Multimodal Approaches

Integrating multimodal approaches in task specification is crucial for enhancing the clarity and precision of robotic task execution. By leveraging visual, linguistic, and gestural data, robots achieve a comprehensive understanding of task requirements, improving performance and adaptability in complex environments. This integration is exemplified by frameworks that extend core formal languages' grammar through naturalization, enabling systems to learn and adapt from user interactions [40].

As illustrated in Figure 3, the integration of multimodal approaches emphasizes key components such as task specification, task planning, and adaptability in dynamic environments. The figure highlights the critical roles of visual, linguistic, and gestural data, as well as the contributions of Large Language Models (LLMs) in simplifying task planning. By identifying and retaining necessary objects for task goals, LLMs streamline planning and reduce computational overhead, thereby enhancing execution efficiency and allowing robots to focus on relevant information while minimizing distractions and errors [30].

Integrating multimodal data fosters adaptable and resilient task specification frameworks capable of responding to dynamic, unstructured environments. This capability is vital for applications requiring the sequencing of unordered events and synthesizing diverse information types, such as in NLP tasks, multimodal instruction summarization, and sequential task planning. Techniques like sequentiality-aware pretraining leverage rich contextual information from various data modalities, improving real-world application performance [41, 42, 43, 34]. By combining linguistic inputs with visual and gestural cues, robots interpret complex instructions more accurately, leading to more effective task execution. This holistic approach underscores multimodal integration's importance in advancing robotic capabilities, ensuring seamless operation in diverse and challenging scenarios.

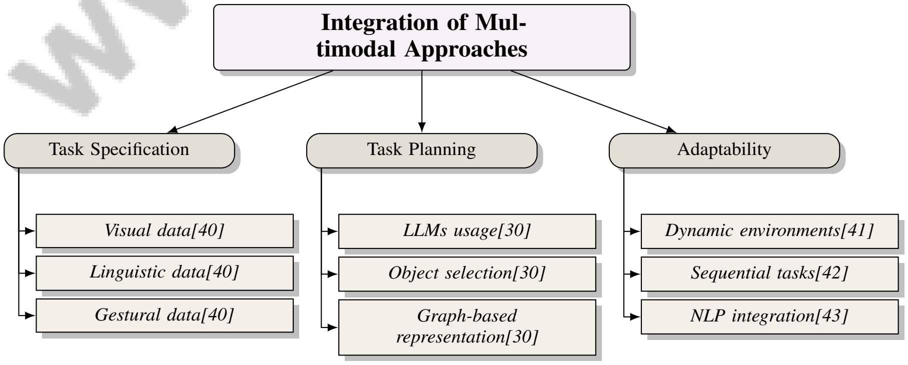


Figure 3: This figure illustrates the integration of multimodal approaches, emphasizing task specification, task planning, and adaptability in dynamic environments. It highlights the use of visual, linguistic, and gestural data, the role of Large Language Models in task planning, and the adaptability required for dynamic environments and sequential tasks.

4 Task Planning and Execution

Category	Feature	Method
Interactive and Adaptive Task Planning	Collaborative Strategies Adaptation and Transfer	GVLM[11] FDM[1]
Graph-Based and Symbolic Methods	Hybrid Reasoning Approaches	DELLT[27], NSAI-RTP[14]
Innovative Task Planning Frameworks	Decomposition Techniques Formal Representation Methods	MLDT[44], TNP[45], LLM-MDTM[46], HIS[39] CTG[47], CRRTP[48]
Integration of Language Models in Task Planning	Self-Improvement Mechanisms Multimodal Integration Adaptive Planning	ISR-LLM[32] ViLaIn[31] GSTL[15], TTG[2]
Adaptive and Real-Time Task Execution	Error Correction Mechanisms	HiCRISP[18], LI[19], HTP-FOON[49]
Collaborative and Multi-Agent Task Planning	Multimodal Understanding Structured Representation Dynamic Strategy Adaptation Communication Enhancement	T&T[4] VG[50] HMTTP[38] PDTPM[17]
Benchmarking and Evaluation of Task Planning Models	Language Model Utilization	RAP[51], LLM-MCTS[29], LLM-TPG[30]

Table 1: This table provides a comprehensive summary of various methods and frameworks used in task planning and execution within robotics. It categorizes these methods into areas such as interactive and adaptive task planning, graph-based and symbolic methods, innovative task planning frameworks, integration of language models, adaptive and real-time task execution, collaborative and multi-agent task planning, and benchmarking and evaluation of task planning models. Each category highlights specific features and methodologies, along with references to relevant research works, illustrating the advancements in robotic adaptability and efficiency.

Task planning and execution are crucial for robotics, facilitating interaction with dynamic environments. Table 1 presents an organized overview of the key methods and frameworks enhancing task planning and execution in robotics, emphasizing their roles in improving adaptability and efficiency in dynamic environments. Additionally, Table 4 offers a comprehensive comparison of various task planning approaches in robotics, emphasizing their adaptability and execution strategies. This section delves into methodologies enhancing robotic adaptability and efficiency, emphasizing interactive and adaptive task planning for real-time navigation and response to environmental shifts.

4.1 Interactive and Adaptive Task Planning

Interactive and adaptive task planning enables robots to effectively navigate and execute tasks in ever-changing environments. The VeriGraph approach uses scene graphs to structure object relationships, allowing for efficient action sequence generation and verification [50]. This framework supports real-time adaptation, ensuring robust task execution. Figure 4 illustrates the key frameworks and methods in interactive and adaptive task planning for robotics, highlighting the VeriGraph approach, GameVLM framework, and multi-robot systems, each contributing to enhanced task execution and adaptability in dynamic environments.

The GameVLM framework enhances decision-making through agent communication and competition, improving accuracy and consistency in dynamic settings [11]. This interaction aids robots in adjusting strategies to maintain high performance.

Multi-robot systems benefit from distributed execution strategies that iteratively refine actions via inter-robot communication [38], optimizing task execution in real-time. The This&That method exemplifies adaptive planning by integrating multimodal inputs like language and gestures, improving interpretation of complex instructions [4].

The Functorial Data Migration method enables task plan transfer across domains, allowing robots to apply learned strategies in new contexts without extensive reprogramming [1]. This adaptability is critical for navigating novel tasks and environments.

Recent advances incorporate multimodal inputs and sequence-aware pretraining, enhancing model adaptability by processing diverse data and adjusting plans in real-time [34]. Advanced techniques ensure robots navigate unpredictable changes while maintaining robust performance and effective human collaboration, with frameworks integrating interactive language models to dynamically generate instructions [20, 52, 53].

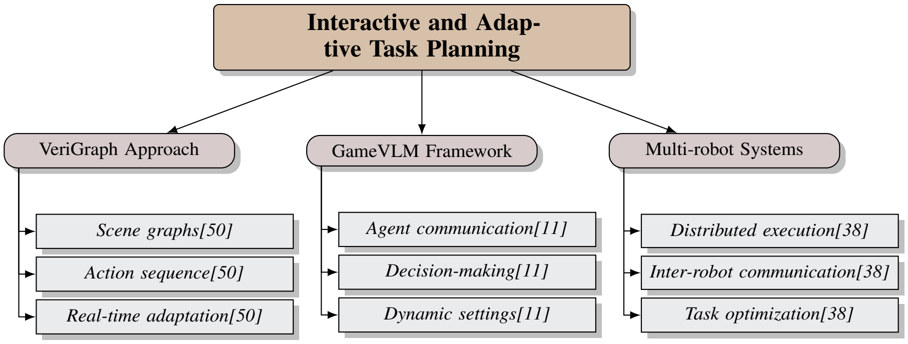


Figure 4: This figure illustrates the key frameworks and methods in interactive and adaptive task planning for robotics, highlighting the VeriGraph approach, GameVLM framework, and multi-robot systems, each contributing to enhanced task execution and adaptability in dynamic environments.

4.2 Graph-Based and Symbolic Methods

Graph-based and symbolic methods provide structured approaches for solving complex planning challenges in robotics. These methods employ advanced graph structures and symbolic representations to model intricate relationships among objects, actions, and goals, facilitating the decomposition of complex tasks into manageable subgoals [45, 54, 55].

Scene graphs enhance spatial and relational information representation, crucial for generating valid action sequences [50]. Symbolic methods use formal languages and logical frameworks, such as Linear Temporal Logic (LTL), to define execution rules and constraints, ensuring systematic execution [27].

The integration of symbolic planning with neural networks allows for the adaptation of task plans based on dynamic constraints, showcasing the synergy between symbolic reasoning and machine learning [14]. Iterative self-refinement of Large Language Models (LLMs) to generate feasible action plans illustrates the potential of combining symbolic methods with language models, enabling robots to refine plans based on feedback [32].

In large-scale domains with partial observations, scalable task planning methods incorporating symbolic reasoning efficiently navigate vast state spaces and generate optimal plans [30].

4.3 Innovative Task Planning Frameworks

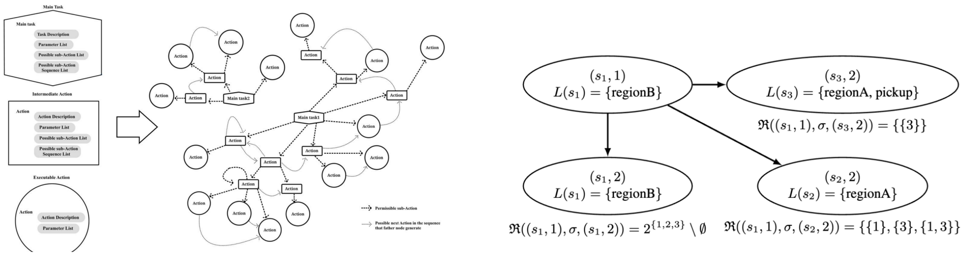
Innovative frameworks in task planning integrate novel methodologies addressing traditional limitations. A significant advancement is formulating Task and Motion Planning (TAMP) problems as hybrid discrete-continuous search problems, ensuring robots generate plans that are both theoretically sound and practically executable [56].

Developments in maneuverability-driven controllers and tool affordance models enhance robotic manipulation in confined spaces, broadening autonomous task scope [46]. Categorical logic for representing world states and actions overcomes classical representation limitations, allowing precise modeling of relationships within planning problems [48].

The Multi-Level Decomposition Technique (MLDT) manages long-horizon tasks by decomposing them into manageable sub-tasks, enhancing open-source LLMs' planning effectiveness [44]. Hierarchical iterative search (HIS) algorithms incrementally evaluate feasible sub-tasks, generating optimal plans with flexibility [39].

Systematic analysis of integration levels between high-level reasoning and low-level checks provides insights into optimizing hybrid planning approaches [19]. This understanding enables robotic systems to efficiently meet strategic objectives and operational constraints.

As depicted in Figure 5, innovative task planning frameworks streamline complex processes and enhance efficiency. The first example outlines task execution into Main Task, Intermediate Actions, and Executable Actions, clarifying task components. The second example illustrates relationships between sets of locations and their labels, emphasizing dynamic interactions in task planning. These


Main Task
Main task
Task Description
Parameter List
Possible sub-Action List
Possible sub-Action Sequence List
Intermediate Action
Action
Action Description
Parameter List
Possible sub-Action List
Possible sub-Action Sequence List
Executable Action
Action
Action Description
Parameter List
Main task1
Main task2
Action
Permissible sub-Action
Possible next Action in the sequence that father node generate
$(s_1, 1)$ $L(s_1) = \{\text{regionB}\}$
$(s_3, 2)$ $L(s_3) = \{\text{regionA, pickup}\}$
$\mathfrak{R}((s_1, 1), \sigma, (s_3, 2)) = \{\{3\}\}$
$(s_1, 2)$ $L(s_1) = \{\text{regionB}\}$
$(s_2, 2)$ $L(s_2) = \{\text{regionA}\}$
$\mathfrak{R}((s_1, 1), \sigma, (s_1, 2)) = 2^{\{1,2,3\}} \setminus \emptyset$
$\mathfrak{R}((s_1, 1), \sigma, (s_2, 2)) = \{\{1\}, \{3\}, \{1, 3\}\}$

(a) A Diagram of a Task Execution Flow[45]

(b) The image shows a directed graph representing a relationship between two sets of locations and their corresponding labels.[47]

Figure 5: Examples of Innovative Task Planning Frameworks

examples underscore the importance of innovative frameworks in effectively managing and executing tasks across diverse operational settings [45, 47].

4.4 Integration of Language Models in Task Planning

Method Name	Integration Techniques	Framework Approaches	Task Planning Adaptability
ViLaIn[31]	Scene Observations	Vision-Language Interpreter	Refine Task Plans
GSTL[15]	Specification Mining Algorithm	Automated Task Planning	Adapt TO Varying
ISR-LLM[32]	Self-refinement Mechanism	Isr-LLM Framework	Iterative Self-refinement
VG[50]	Scene Graphs	Verigraph Framework	Iterative Verification Mechanism
PDTPM[17]	Prompt-driven Control	Prompt-driven Method	Real-time Adjustments
TTG[2]	Semantic Similarity	Task Tree Generation	Dynamic Generation
HTP-FOON[49]	Two-level Planning	Hierarchical Planning Approach	Adapting TO Scenarios
T			
	T[4]	Video Diffusion Model	This&that
Multimodal Inputs			

Table 2: Comparison of various frameworks integrating language models into robotic task planning, highlighting their respective integration techniques, framework approaches, and adaptability in task planning. The table provides an overview of eight distinct methods, demonstrating the diverse strategies employed to enhance decision-making and execution precision in robotic systems.

Integrating language models into task planning has significantly advanced robotic decision-making, adaptability, and contextual understanding. Table 2 presents a comprehensive comparison of state-of-the-art frameworks that integrate language models into robotic task planning, showcasing the innovative techniques and approaches utilized to improve adaptability and contextual understanding. State-of-the-art Large Language Models (LLMs) and vision-language models, such as in the ViLaIn framework, enhance problem descriptions, improving execution precision [31]. The automated task planning framework utilizing Generalized Spatio-Temporal Logic (GSTL) captures spatial and temporal information, facilitating comprehensive task planning [15].

The ISR-LLM framework introduces a self-refinement mechanism for LLMs to iteratively correct generated plans, aligning execution with objectives [32]. The CLIMB framework employs a hybrid neuro-symbolic approach for incremental model building and self-improvement through environmental interaction, crucial for continuous learning in dynamic settings [16].

Vision-language models are integrated into task planning through frameworks like VeriGraph, which utilize scene graphs to verify action sequence feasibility, ensuring contextually and spatially coherent plans [50]. The prompt-driven task planning method leverages language models for drones to execute tasks based on natural language commands, demonstrating their versatility in robotic applications [17].

The Task Tree Generation method adapts knowledge from the Functional Object-Oriented Network (FOON) to new contexts using semantic similarity, enhancing task planning adaptability [2]. Transforming FOON into Planning Domain Definition Language (PDDL) facilitates executable task plan generation through off-the-shelf planners, highlighting the integration of language models with traditional frameworks [49].

Additionally, a benchmark by Agrawal et al. evaluates commonsense reasoning through utility, context, and physical state, underscoring language models' role in enhancing task planning through intuitive understanding [3]. The This&That method illustrates language model integration by leveraging multimodal inputs for clearer task communication and reduced execution ambiguity [4].

4.5 Adaptive and Real-Time Task Execution

Adaptive and real-time task execution is essential for robotic systems to navigate and operate in dynamic environments. The HiCRISP hierarchical structure exemplifies real-time error correction capabilities, significantly enhancing task execution success rates by allowing robots to respond to real-time feedback [18]. This adaptability is vital for efficient task execution in unforeseen scenarios.

Integrating low-level reasoning during planning, as demonstrated by Erdem et al., allows for dynamic feasibility checks, ensuring viable actions are pursued and enhancing plan quality [19]. Frameworks supporting real-time adaptability enable robots to maintain high performance in complex and unpredictable environments.

The methodology proposed by Paulius et al. emphasizes flexibility in adapting to new scenarios, achieving significant reductions in planning time complexity compared to traditional methods [49]. This flexibility is crucial for managing long-horizon tasks and ensuring robots can adjust plans in response to evolving conditions.

By integrating advanced frameworks and methodologies, such as relevance frameworks for scene understanding and adaptive task planning strategies, robotic systems enhance adaptability and responsiveness. They intelligently filter and process relevant information, collaborate effectively with humans by aligning task allocation with preferences, and leverage large language models for seamless communication and decentralized execution among heterogeneous robots. These capabilities significantly improve task execution efficiency across various scenarios, ensuring successful outcomes in human-robot collaboration [20, 21, 57]. Real-time decision-making and environmental adaptation are crucial for advancing robotic systems in real-world applications.

4.6 Collaborative and Multi-Agent Task Planning

Collaborative and multi-agent task planning is vital for enhancing robotic systems' efficiency in complex environments. Coordination among agents is essential for achieving collective goals, facilitated by robust communication protocols and shared environmental representations. The VeriGraph approach integrates scene graphs, enabling agents to share and verify environmental information, promoting a cohesive understanding and improving collaborative planning [50].

The hierarchical framework for multi-robot systems synthesizes optimized task execution strategies, highlighting the importance of reducing wait times during collaboration to enhance overall efficiency [38]. This framework allows robots to dynamically adjust strategies based on real-time feedback, maintaining high performance in collaborative scenarios.

Incorporating language models into multi-agent planning enhances communication and coordination. The prompt-driven task planning method enables drones to execute tasks based on natural language commands, showcasing language models' potential to streamline communication and improve multi-agent adaptability [17]. This underscores the value of multimodal inputs, as demonstrated by the This&That method, which allows agents to interpret complex instructions through language and gestures, reducing ambiguity in task execution [4].

Frameworks supporting distributed execution strategy adjustment emphasize inter-agent communication's importance in dynamically adapting strategies [38]. This capability is crucial for multi-agent systems to adapt to changing environments and maintain high performance levels.

4.7 Benchmarking and Evaluation of Task Planning Models

Benchmarking and evaluating task planning models are critical for assessing robotic systems' performance in executing complex tasks. Evaluation typically involves metrics measuring aspects such as success rates, planning efficiency, and adaptability to dynamic environments, reflecting the time taken to solve planning problems, essential for real-time applications in robotics [58]. Table 3 provides a

Benchmark	Size	Domain	Task Format	Metric
PDDL-ASP[58]	3	Robotics	Task Planning	Planning Time
TEACh[59]	1,000	Household Task Planning	Execution From Dialog History (edh)	Success Rate, Edit Distance
Transformers[60]	2,097	Text Classification	Question Answering	Accuracy, F1-score
OSSA[61]	184	Robotics	Instruction Following	Accuracy, Completion Accuracy
COST[62]	20,000	Task Planning	Action Steps Generation	Success Rate
SheetCopilot[63]	221	Spreadsheet Manipulation	Task Execution	Exec@1, Pass@1
InterAct[64]	3,000	Task Planning	Multi-step Task Execution	Success Rate
VIMA-BENCH[65]	650,000	Robot Manipulation	Multimodal Prompting	Success Rate

Table 3: This table summarizes representative benchmarks utilized in the evaluation of task planning models across various domains. It includes key details such as the size of each benchmark, the specific domain it pertains to, the task format, and the metrics used for evaluation. These benchmarks provide a comprehensive overview of the diverse methodologies and evaluation criteria employed in the field of task planning.

detailed overview of the benchmarks used in assessing task planning models, highlighting their size, domain, task format, and evaluation metrics.

As illustrated in Figure 6, this figure presents a hierarchical classification of task planning models, focusing on evaluation metrics, frameworks and algorithms, and comparative analysis. It highlights key evaluation criteria such as success rates, planning efficiency, and adaptability, and provides an overview of prominent frameworks like RAP, LLM-MCTS, and LLM-TPG. Additionally, it compares different planning methodologies, including PDDL vs ASP and RAP vs baseline models, offering insights into their performance and applicability.

The Reasoning and Planning (RAP) framework demonstrates substantial improvements over existing methods, achieving higher success rates in plan generation and reasoning tasks, highlighting the importance of specialized frameworks in enhancing capabilities [51]. RAP's success underscores the need for benchmarking against advanced reasoning systems to ensure robustness.

The performance of Large Language Model-based Monte Carlo Tree Search (LLM-MCTS) was assessed by comparing success rates in task completion against baseline methods, focusing on its ability to reorganize household items based on natural language instructions [29]. This comparison reveals the advantages of integrating language models into task planning.

Furthermore, the scalability and efficiency of LLM-based planners were evaluated against traditional methods like Monte Carlo Tree Search (MCTS) in reduced state spaces, emphasizing their potential to improve task planning in complex domains [30]. Evaluating these models in reduced state spaces is crucial for understanding their scalability and adaptability.

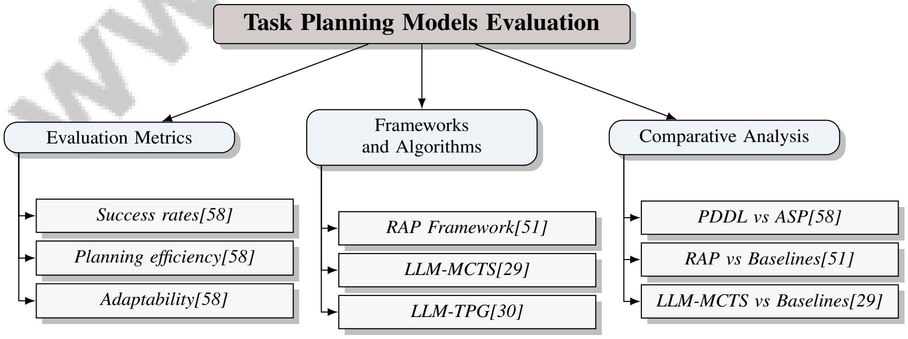


Figure 6: This figure presents a hierarchical classification of task planning models, focusing on evaluation metrics, frameworks and algorithms, and comparative analysis. It highlights key evaluation criteria such as success rates, planning efficiency, and adaptability, and provides an overview of prominent frameworks like RAP, LLM-MCTS, and LLM-TPG. Additionally, it compares different planning methodologies, including PDDL vs ASP and RAP vs baseline models, offering insights into their performance and applicability.

Feature	Interactive and Adaptive Task Planning	Graph-Based and Symbolic Methods	Innovative Task Planning Frameworks
Adaptability	Real-time Adaptation	Dynamic Constraints	Hybrid Search
Execution Strategy	Agent Communication	Symbolic Planning	Hierarchical Decomposition
Integration Method	Scene Graphs	Scene Graphs	Categorical Logic

Table 4: This table provides a comparative analysis of three distinct task planning methodologies in robotics: Interactive and Adaptive Task Planning, Graph-Based and Symbolic Methods, and Innovative Task Planning Frameworks. It highlights key features such as adaptability, execution strategy, and integration method, showcasing the diverse approaches employed to enhance robotic task execution in dynamic environments.

5 Natural Language Processing and Language Models

The integration of large language models (LLMs) with knowledge-based systems marks a pivotal advancement in natural language processing (NLP), significantly enhancing robotic systems' capabilities in executing complex tasks. This convergence facilitates the extraction and organization of information from unstructured text, enabling precise action plans based on natural language instructions while addressing multimodal interaction challenges. Frameworks incorporating visual perception with language comprehension demonstrate LLMs' potential in embodied tasks, enriching human-robot interactions and expanding robotics applications [36, 66, 13, 35, 41]. By harnessing linguistic capabilities and structured knowledge, these integrations enhance decision-making processes. This section explores mechanisms and frameworks exemplifying this integration, focusing on the effective combination of language models and knowledge-based systems to improve task planning and execution.

5.1 Integration of Language Models with Knowledge-Based Systems

Integrating language models with knowledge-based systems enhances task planning in robotics by leveraging language models and structured repositories for context-aware decision-making. The Wonderful Team project exemplifies this by using a Vision-Large Language Model framework, enabling robots to perform tasks without prior specific training [67]. Transformer-based models, as highlighted by Wolf et al., provide a robust foundation for integrating pretrained models across domains, facilitating seamless deployment and adaptability in dynamic environments [60]. The SayComply project underscores the importance of grounding robotic actions in real-world contexts by incorporating operational knowledge into task planning [68].

Vision Language Models (VLMs) are pivotal in this integration, enhancing user interaction by collecting on-site information, as seen in remote life support robots [69]. The V I L A framework supports multimodal goal specifications and dynamic task execution, enabling robots to adapt to changing conditions and user requirements [70]. Integrating transformer language models with knowledge-based systems is crucial for improving NLP applications and task planning by incorporating domain-specific knowledge into decision-making processes [42]. This integration benefits requirements engineering, where NLP automates the generation of Unified Modeling Language (UML) diagrams, streamlining the design of complex systems [71].

5.2 Advancements in Vision-Language Models

Recent advancements in vision-language models (VLMs) have significantly enhanced robotic systems, enabling sophisticated interactions and task executions by integrating visual perception with NLP for better contextual understanding and decision-making [11]. A key advancement is processing multimodal inputs, essential for tasks requiring understanding of both visual and linguistic information. Frameworks using scene graphs to represent relationships between objects and actions exemplify a structured approach to task execution based on visual and language cues [50]. By leveraging these models, robots achieve a comprehensive understanding of their environment, leading to accurate and efficient task execution.

Integrating VLMs into task planning processes has facilitated adaptable and responsive robotic systems. The VeriGraph approach allows for verifying action sequences based on visual and language inputs, ensuring contextually coherent and feasible task plans [50]. This integration enhances robots' adaptability to dynamic environments and precision in task execution. Furthermore, advancements in VLMs support real-time interaction and adaptation, allowing robots to adjust actions based on real-time feedback, crucial for maintaining high performance in complex scenarios [11].

5.3 Challenges in Generating Factual and Consistent Responses

Generating factual and consistent responses in knowledge-intensive tasks poses significant challenges for language models, particularly LLMs. A primary issue is hallucination, where models generate plausible but ungrounded information, often exacerbated by difficulty in acquiring long-tailed knowledge, necessary for accurate responses [72]. Another challenge is limited memory expansion capacity, restricting models' ability to retain extensive knowledge over time, hindering consistent responses across contexts and tasks [72].

Evaluating model performance is often constrained by existing benchmarks, which may not comprehensively cover diverse tasks and models, complicating consistent assessment and comparison of factual accuracy and consistency [60]. There is a need for robust evaluation frameworks capturing the nuances of factual consistency in language model outputs. In spoken instruction contexts, the quality of automatic speech recognition (ASR) output significantly impacts response consistency and accuracy. Methods like SIFToM, relying on ASR, may produce errors under challenging conditions, leading to inconsistencies in interpreting spoken commands [73]. Addressing these challenges requires advancements in both language modeling and speech recognition technologies to ensure factually accurate, contextually appropriate, and consistent responses.

5.4 Role of Language Models in Task Planning and Execution

Language models are pivotal in enhancing task planning and execution by providing sophisticated mechanisms for interpreting and generating natural language instructions, improving human-robot interaction. Integrating LLMs within task planning frameworks significantly augments robotic systems' accuracy and reliability. In the HERACLEs framework, language models generate action plans aligned with user-friendly task specifications from natural language, facilitating intuitive task execution [74]. Feedback mechanisms within LLMs, as highlighted by Huang et al., allow models to adapt plans based on real-time observations, ensuring responsive execution to environmental changes [75].

This adaptability is further enhanced by multimodal LLMs, enabling robots to perform complex tasks with improved efficiency by leveraging diverse data inputs [36]. Language models' importance in task planning is underscored by their role in Vision-Language-Action (VLA) models, integrating visual perception with language processing for effective task execution [24]. Shirai et al.'s approach advances the integration of language-guided planning with symbolic methods, ensuring plans are contextually relevant and symbolically coherent [31].

In dynamic interactions between decision agents, integrating VLMs with a zero-sum game strategy exemplifies language models' role in enhancing task planning, allowing dynamic strategy adjustments based on real-time interactions, improving decision-making [11]. Despite advancements, challenges remain in navigating complex scenarios involving multiple contextual factors and physical states, as demonstrated by benchmarks assessing large language models' reasoning capabilities. While these models exhibit strong reasoning in object selection, handling intricate scenarios is still developing [3].

5.5 Multimodal and Multilingual Capabilities

Incorporating multimodal and multilingual capabilities in language models is crucial for advancing robotic systems, allowing more nuanced and effective human-robot interactions. Multimodal capabilities enable robots to process diverse data types, such as visual, auditory, and textual inputs, enhancing their ability to execute complex tasks in dynamic environments [24]. This integration is exemplified by VLMs, combining visual perception with language processing for comprehensive task execution [11].

Processing multimodal inputs is critical for robots in real-world scenarios, where they must interpret and respond to various sensory data. The VeriGraph approach utilizes scene graphs to represent spatial and relational information, enabling robots to verify action sequences based on visual and

AI-generated, for reference only.

language cues [50]. This ensures task plans are contextually coherent and feasible, allowing robots to adapt to changing conditions and execute tasks with precision.

Multilingual capabilities further enhance language models' versatility in robotics, enabling robots to interact with users across different linguistic contexts, crucial for global applications where communication with diverse languages is necessary. Integrating multilingual models allows robots to interpret and generate instructions in multiple languages, facilitating inclusive and accessible human-robot interactions [36]. The necessity for multilingual capabilities is highlighted by robots' requirement to function efficiently across various cultural and linguistic contexts, as effective communication is critical for tasks such as conducting research and interacting with diverse user populations [41, 76]. By leveraging multilingual language models, robotic systems can overcome language barriers and provide more personalized, user-friendly interactions, enhancing their functionality and appeal in international markets.

6 Human-Robot Interaction

6.1 Collaborative Task Execution

Collaborative task execution is fundamental to enhancing the synergy and efficiency of human-robot teams, requiring robots to adapt to human preferences and modify their actions based on real-time feedback. The lead-follow adaptive robot approach exemplifies this adaptability, aligning robotic actions with human preferences to improve team performance and positively influence human perceptions [20]. The Directed Acyclic Graph Plan (DAG-Plan) framework optimizes collaboration by dynamically assigning tasks to robotic arms, enhancing execution efficiency and enabling real-time adaptability to environmental changes [77].

The Epistemic Human-Aware Task Planner (EHATP) framework advances collaboration by anticipating human needs, allowing robots to proactively adjust their plans for smoother and more efficient task execution [78]. This anticipatory capability enables seamless integration of robots into human workflows, enhancing team performance. The Tool-Planner framework addresses tool usage challenges by dynamically adjusting plans in response to tool errors, significantly boosting efficiency and fostering effective collaboration [79].

Integrating knowledge representation with task planning improves collaborative execution by enabling skill transfer across different robotic systems, enhancing flexibility and adaptability [7]. Leveraging shared knowledge allows robots to better understand and respond to human actions, facilitating seamless interactions. Goal inference is also crucial for improving collaboration efficiency and user satisfaction, as it enables robots to anticipate and adapt to human actions, aligning their objectives with those of their human counterparts [80]. This alignment is vital for ensuring effective human-robot teams achieve shared goals with minimal friction.

7 Challenges and Future Directions

In robotics, addressing challenges and future directions is essential for fostering innovation and enhancing system capabilities. Technological advancements significantly influence the evolution of robotic systems, particularly in task execution, adaptability, and interaction.

7.1 Technological Advancements and Future Directions

The future of robotics is shaped by advancements aimed at overcoming challenges in task execution, adaptability, and interaction. Integrating symbolic and neural components enhances learning in dynamic environments and comprehension of social norms [14], promising more intelligent robots capable of navigating complex scenarios. In task specification and planning, Large Language Models (LLMs) within Business Process Management (BPM) open new research avenues for process automation and decision-making [13]. Future research will enhance LLM's autonomy in applying LLM+P, reducing human input reliance [26].

Frameworks like ViLaIn improve interpretability and logical correctness in robot planning, enhancing transparency and trust [31]. Incorporating motion planning in frameworks such as ISR-LLM is expected to broaden applicability, improving reliability in task and motion planning [32]. Future

AI-generated, for reference only.

directions emphasize refining interaction protocols and expanding function libraries for multi-drone systems, boosting operational capabilities [17]. Advances in scene graph generation methods will enhance accuracy in complex environments, facilitating better task execution [50].

Privacy and security remain paramount, with solutions focusing on strong privacy guarantees alongside high model accuracy [12]. Robust predicate grounding and refining hybrid approaches are vital for enhancing adaptability across diverse environments [16]. Future research should enhance perception modules and conduct comprehensive evaluations of system properties in varied environments [18]. The EgoPlan-Bench indicates that current MLLMs lack mature human-level planning capabilities, underscoring the need for further advancements [8]. Future work will focus on automating recipe generation, enriching the FOON knowledge base, and executing generated task trees for practical evaluation [2].

Research could explore hybrid approaches combining precomputation with replanning strategies to enhance planning efficiency [19]. Integrating re-planning capabilities and improving collision avoidance mechanisms will bolster robustness in dynamic environments [49]. Directions also include enhancing the expressivity of translation functors and adapting frameworks for partially ordered plans, addressing current task planning challenges [1].

Benchmarks reveal significant gaps in AI's ability to sequence multimodal instructions compared to human performance, highlighting the need for improved models in multimodal reasoning [34]. Future research will extend the HIS approach to multi-agent systems and explore optimizations for scalability in dynamic environments [39]. Enhancing gesture recognition robustness and extending the This&That method to complex tasks indicate potential advancements in robotics [4].

7.2 Handling Real-World Complexities

Addressing real-world complexities in task planning and execution presents significant challenges due to the dynamic nature of such environments. Inefficient wait times during collaborations, caused by synchronization constraints, lead to suboptimal performance, particularly in multi-robot systems [38]. Accurate interpretation of object relationships and constraints is another obstacle, often resulting in incorrect action sequences [50]. This challenge is exacerbated by the need to manage unforeseen situations and inconsistencies in outputs from multiple agents [11].

Limitations in grounding continuous state observations to logical predicates complicate handling complex environments [16], highlighting the necessity for robust frameworks bridging high-level reasoning and low-level execution. The reliance on idealized scenarios in benchmarks may not fully capture real-world complexities, limiting their applicability in task planning and execution [3]. Accurate representation of workspaces and defined soft constraints impact the feasibility of generated plans [39], necessitating precise modeling.

Handling real-world complexities is supported by methods adapting to dynamic environments without extensive retraining, crucial for maintaining flexibility and responsiveness [17]. However, inaccuracies in ingredient substitutions and contextual relevance of states remain significant challenges [2]. Finally, the limitations of open-loop execution, which may not effectively manage unexpected contingencies, underscore the need for adaptive planning frameworks [49].

7.3 Improving Perception and Interaction

Enhancing robot perception and interaction capabilities is fundamental for advancing human-robot collaboration in complex environments. Structured help feedback allows AI systems to seek clarification and improve task understanding [81], enhancing interaction by enabling robots to ask for additional information, reducing errors and improving accuracy. Integrating privacy-preserving techniques in machine learning enhances user trust and data security, critical for improving interaction capabilities in mobile robotic systems [12].

Future research should refine the object selection process and explore further integrations of LLMs with advanced planning algorithms to enhance performance in complex scenarios [30]. Developing sophisticated models that understand task dependencies and step interchangeability will improve adaptability to dynamic environments [34]. The accuracy of scene graph generation, often suffering from occlusion and poor performance in complex scenes, requires advancements in perception technologies [50]. Moreover, the need for carefully crafted prompts and predefined action constraints limits flexibility, highlighting the importance of developing adaptable interaction frameworks [36].

7.4 Enhancing Language Model Capabilities

Enhancing language model capabilities is pivotal for advancing task planning and execution, particularly in dynamic environments. Integrating advanced reasoning techniques and task decomposition improves low-parameter LLM functionality, addressing planning and execution limitations [82]. Frameworks like PlanSys2 enhance language model capabilities, essential for improving planning in real-world applications [83]. Future work should explore integrating advanced NLP methodologies and developing better-labeled datasets to enhance NLP applications [33]. Exploring improved methods for tool selection within clusters can enhance language model capabilities, facilitating effective task planning [79].

Improving language models' robustness in handling multiple LTL structures and translating ambiguous commands is critical [27], including using reinforcement learning-based fine-tuning methods, particularly in multi-robot systems [84]. Addressing timing and uncertainty complexities requires enhancements in language model capabilities [25]. Future directions include improving predictive capabilities for long-term task planning [11].

Enhancing Multimodal Large Language Models (MLLMs) by integrating additional sensory modalities and improving contextual understanding can significantly advance planning capabilities [8]. Enhancing language models to improve object detection robustness in dynamic environments ensures effective operation in real-world scenarios [15].

7.5 Adapting to Dynamic Environments

Adaptability to dynamic environments is crucial for robotic systems, enabling effective responses to changes during task execution. This adaptability is vital for maintaining high performance and reliability in real-world scenarios. Frameworks supporting introspective and extrospective dialogue, like MultiTalk, enhance adaptability by enabling systems to engage in complex dialogues and generate approximate object models in open-set simulations [85].

Understanding human decision-making processes is essential for robots collaborating with humans. Future research should refine models of human decision-making and explore realistic control of human avatars to enhance robot adaptability and improve collaboration effectiveness [80]. By understanding human behavior, robots can anticipate changes and adjust actions accordingly, leading to smoother task execution.

Advancements in sensor technologies and data processing capabilities support adaptability by allowing robots to perceive and interpret environmental changes accurately. Enhanced perception empowers robots to make informed decisions and adapt strategies in real-time. Utilizing advanced frameworks, such as the relevance framework for scene understanding and integrating human expertise into task planning, can significantly improve collaborative performance. Innovative approaches like Robotic Vision-Language Planning (ViLa) enable robots to incorporate perceptual data and commonsense knowledge into planning, enhancing responsiveness to changing conditions, fostering effective human-robot collaboration [21, 53, 70, 86, 87].

7.6 Bridging the Gap in Task Generalization

Addressing task generalization challenges is essential for improving adaptability and versatility, enabling effective operation across various environments. Recent advancements in mapping natural language task specifications to LTL enhance robots' ability to understand and execute complex commands, even with unseen object references [88]. Representation pretraining approaches allow robots to plan multi-step manipulation tasks in novel situations, leveraging semantic and geometric information from large-scale datasets, improving success rates in household activities [89]. Task generalization enables robots to apply learned knowledge to new situations without extensive retraining, critical for autonomous systems in unpredictable settings.

Integrating multimodal learning approaches enables robots to process and synthesize information from various sensory inputs, enhancing task generalization [24]. Hierarchical task representations break down complex tasks into manageable sub-tasks, improving learning efficiency and enabling strategy adaptation to different contexts [77]. Incorporating reinforcement learning techniques allows robots to learn from environmental interactions, continuously refining strategies based on feedback, promoting flexible behaviors across various tasks [82].

Developing robust transfer learning frameworks is critical for enhancing task generalization, enabling robots to leverage existing knowledge from related tasks to accelerate learning in new situations, reducing the need for extensive data collection and retraining [26]. Integrating advanced language models, such as LLMs, with task planning frameworks can significantly improve task generalization, facilitating the interpretation and generation of natural language instructions, enhancing knowledge transfer across tasks and environments [36].

8 Conclusion

This survey highlights the pivotal role of interdisciplinary methodologies in advancing task specification, planning, and execution within robotics. The confluence of robotics, artificial intelligence, and natural language processing has significantly refined the rationality and success rates of task planning. Notable systems such as RoboGPT exemplify these advancements by enhancing the capabilities of intelligent agents. The PsALM framework underscores the importance of interdisciplinary strategies in crafting precise and reliable mission designs, thereby fostering the development of intelligent and adaptable robotic systems.

Frameworks like RAHL showcase the benefits of integrating hierarchical reinforcement learning with retrieval-augmented techniques, thereby improving decision-making processes in robotic tasks. Similarly, the incorporation of Graph Neural Networks into task planning illustrates the potential of graph learning methods to elevate language agent performance across various benchmarks, underscoring the value of diverse computational models in enhancing robotic adaptability and effectiveness in dynamic settings.

In multi-human single-robot interactions, collaborative frameworks designed to enhance team performance have demonstrated their efficacy in simulated tasks, highlighting the necessity of optimizing human-robot interactions for efficient task execution. Moreover, the application of Large Language Models in tool-object manipulation, when combined with maneuverability-driven strategies, marks a significant leap in robotic capabilities. The PSM* method exemplifies advancements in reducing path length and computation time, thereby boosting robot autonomy in practical scenarios.

The survey also emphasizes the robust generalization capabilities of Neural Task Programming across different task structures, underscoring the need for innovative approaches in robotics. The PROGPROMPT method illustrates the effectiveness of structured programming prompts in task planning, significantly outperforming existing techniques in generating executable task plans. Additionally, the SkiROS2 platform demonstrates how a skill-based robot control system can adeptly address the complexities of modern autonomous tasks by integrating various robotic systems, task-level planning, and user learning.

References

[1] Angeline Aguinaldo, Evan Patterson, and William Regli. Automating transfer of robot task plans using functorial data migrations, 2024.

[2] Md. Sadman Sakib, David Paulius, and Yu Sun. Approximate task tree retrieval in a knowledge network for robotic cooking, 2022.

[3] Ayush Agrawal, Raghav Prabhakar, Anirudh Goyal, and Dianbo Liu. Physical reasoning and object planning for household embodied agents, 2024.

[4] Boyang Wang, Nikhil Sridhar, Chao Feng, Mark Van der Merwe, Adam Fishman, Nima Fazeli, and Jeong Joon Park. Thisthat: Language-gesture controlled video generation for robot planning, 2024.

[5] Xiaohan Zhang, Zainab Altaweel, Yohei Hayamizu, Yan Ding, Saeid Amiri, Hao Yang, Andy Kaminski, Chad Esselink, and Shiqi Zhang. Dkprompt: Domain knowledge prompting vision-language models for open-world planning, 2024.

[6] 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.

[7] Matthias Mayr, Faseeh Ahmad, Alexander Duerr, and Volker Krueger. Using knowledge representation and task planning for robot-agnostic skills on the example of contact-rich wiping tasks, 2023.

[8] Yi Chen, Yuying Ge, Yixiao Ge, Mingyu Ding, Bohao Li, Rui Wang, Ruifeng Xu, Ying Shan, and Xihui Liu. Egoplan-bench: Benchmarking multimodal large language models for human-level planning, 2024.

[9] Ali Noormohammadi-Asl, Kevin Fan, Stephen L. Smith, and Kerstin Dautenhahn. Human leading or following preferences: Effects on human perception of the robot and the human-robot collaboration, 2024.

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

[11] Aoran Mei, Jianhua Wang, Guo-Niu Zhu, and Zhongxue Gan. Gamevlm: A decision-making framework for robotic task planning based on visual language models and zero-sum games, 2024.

[12] Felix Burget, Lukas Dominique Josef Fiederer, Daniel Kuhner, Martin Völker, Johannes Aldinger, Robin Tibor Schirrmeister, Chau Do, Joschka Boedecker, Bernhard Nebel, Tonio Ball, and Wolfram Burgard. Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills, 2018.

[13] Michael Grohs, Luka Abb, Nourhan Elsayed, and Jana-Rebecca Rehse. Large language models can accomplish business process management tasks, 2023.

[14] Niklas Hemken, Florian Jacob, Fabian Peller-Konrad, Rainer Kartmann, Tamim Asfour, and Hannes Hartenstein. How to raise a robot – a case for neuro-symbolic ai in constrained task planning for humanoid assistive robots, 2023.

[15] Zhiyu Liu, Meng Jiang, and Hai Lin. Specification mining and automated task planning for autonomous robots based on a graph-based spatial temporal logic, 2020.

[16] Walker Byrnes, Miroslav Bogdanovic, Avi Balakirsky, Stephen Balakirsky, and Animesh Garg. Climb: Language-guided continual learning for task planning with iterative model building, 2024.

[17] Yaohua Liu. A prompt-driven task planning method for multi-drones based on large language model, 2024.

[18] Chenlin Ming, Jiacheng Lin, Pangkit Fong, Han Wang, Xiaoming Duan, and Jianping He. Hicrisp: An llm-based hierarchical closed-loop robotic intelligent self-correction planner, 2024.

[19] Esra Erdem, Volkan Patoglu, and Peter Schüller. Levels of integration between low-level reasoning and task planning, 2013.

[20] Ali Noormohammadi-Asl, Stephen L. Smith, and Kerstin Dautenhahn. To lead or to follow? adaptive robot task planning in human-robot collaboration, 2024.

[21] Xiao-Tong Zhang, Ding-Cheng Huang, and Kamal Youcef-Toumi. Relevance for human robot collaboration, 2024.

[22] Jonathan Francis, Nariaki Kitamura, Felix Labelle, Xiaopeng Lu, Ingrid Navarro, and Jean Oh. Core challenges in embodied vision-language planning, 2023.

[23] Isabel M. Rayas Fernández. Advancing robot autonomy for long-horizon tasks, 2023.

[24] Yueen Ma, Zixing Song, Yuzheng Zhuang, Jianye Hao, and Irwin King. A survey on vision-language-action models for embodied ai, 2024.

[25] Till Hofmann. Towards bridging the gap between high-level reasoning and execution on robots, 2023.

[26] Bo Liu, Yuqian Jiang, Xiaohan Zhang, Qiang Liu, Shiqi Zhang, Joydeep Biswas, and Peter Stone. Llm+ p: Empowering large language models with optimal planning proficiency. *arXiv preprint arXiv:2304.11477*, 2023.

[27] Jiayi Pan, Glen Chou, and Dmitry Berenson. Data-efficient learning of natural language to linear temporal logic translators for robot task specification. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11554–11561. IEEE, 2023.

[28] Muhayy Ud Din, Jan Rosell, Waseem Akram, Isiah Zaplana, Maximo A Roa, Lakmal Seneviratne, and Irfan Hussain. Ontology-driven prompt tuning for llm-based task and motion planning, 2024.

[29] Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning, 2023.

[30] Rodrigo Pérez-Dattari, Zhaoting Li, Robert Babuška, Jens Kober, and Cosimo Della Santina. Scalable task planning via large language models and structured world representations, 2025.

[31] Keisuke Shirai, Cristian C. Beltran-Hernandez, Masashi Hamaya, Atsushi Hashimoto, Shohei Tanaka, Kento Kawaharazuka, Kazutoshi Tanaka, Yoshitaka Ushiku, and Shinsuke Mori. Vision-language interpreter for robot task planning, 2024.

[32] Zhehua Zhou, Jiayang Song, Kunpeng Yao, Zhan Shu, and Lei Ma. Isr-llm: Iterative self-refined large language model for long-horizon sequential task planning, 2023.

[33] Yue Kang, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. Natural language processing (nlp) in management research: A literature review. *Journal of Management Analytics*, 7(2):139–172, 2020.

[34] Te-Lin Wu, Alex Spangher, Pegah Alipoormolabashi, Marjorie Freedman, Ralph Weischedel, and Nanyun Peng. Understanding multimodal procedural knowledge by sequencing multimodal instructional manuals, 2024.

[35] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56(2):1–40, 2023.

[36] Jiaqi Wang, Zihao Wu, Yiwei Li, Hanqi Jiang, Peng Shu, Enze Shi, Huawen Hu, Chong Ma, Yiheng Liu, Xuhui Wang, Yincheng Yao, Xuan Liu, Huaqin Zhao, Zhengliang Liu, Haixing Dai, Lin Zhao, Bao Ge, Xiang Li, Tianming Liu, and Shu Zhang. Large language models for robotics: Opportunities, challenges, and perspectives, 2024.

[37] Nathan Dolbir, Triyasha Dastidar, and Kaushik Roy. Nlp is not enough – contextualization of user input in chatbots, 2021.

[38] Ruofei Bai, Ronghao Zheng, Yang Xu, Meiqin Liu, and Senlin Zhang. Hierarchical multi-robot strategies synthesis and optimization under individual and collaborative temporal logic specifications, 2021.

[39] Ziyang Chen, Zhangli Zhou, Lin Li, and Zhen Kan. Soft task planning with hierarchical temporal logic specifications. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2299–2304. IEEE, 2024.

[40] Ivan Gavran, Brendon Boldt, Eva Darulova, and Rupak Majumdar. Precise but natural specification for robot tasks, 2018.

[41] Hao Kang and Chenyan Xiong. Researcharena: Benchmarking large language models' ability to collect and organize information as research agents, 2025.

[42] Ross Gruetzemacher and David Paradice. Deep transfer learning beyond: Transformer language models in information systems research, 2021.

[43] Jonas Becker, Jan Philip Wahle, Bela Gipp, and Terry Ruas. Text generation: A systematic literature review of tasks, evaluation, and challenges, 2024.

[44] Yike Wu, Jiatao Zhang, Nan Hu, LanLing Tang, Guilin Qi, Jun Shao, Jie Ren, and Wei Song. Mldt: Multi-level decomposition for complex long-horizon robotic task planning with open-source large language model, 2024.

[45] Yue Zhen, Sheng Bi, Lu Xing-tong, Pan Wei-qin, Shi Hai-peng, Chen Zi-rui, and Fang Yi-shu. Robot task planning based on large language model representing knowledge with directed graph structures, 2023.

[46] Hoi-Yin Lee, Peng Zhou, Anqing Duan, Wanyu Ma, Chenguang Yang, and David Navarro-Alarcon. Non-prehensile tool-object manipulation by integrating llm-based planning and manoeuvrability-driven controls, 2025.

[47] Amy Fang and Hadas Kress-Gazit. High-level, collaborative task planning grammar and execution for heterogeneous agents, 2024.

[48] Angeline Aguinaldo, Evan Patterson, James Fairbanks, William Regli, and Jaime Ruiz. A categorical representation language and computational system for knowledge-based planning, 2023.

[49] David Paulius, Alejandro Agostini, and Dongheui Lee. Long-horizon planning and execution with functional object-oriented networks, 2023.

[50] Daniel Ekpo, Mara Levy, Saksham Suri, Chuong Huynh, and Abhinav Shrivastava. Verigraph: Scene graphs for execution verifiable robot planning, 2024.

[51] Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. *arXiv preprint arXiv:2305.14992*, 2023.

[52] Boyi Li, Philipp Wu, Pieter Abbeel, and Jitendra Malik. Interactive task planning with language models, 2025.

[53] Abhinav Dahiya and Stephen L. Smith. Adaptive robot assistance: Expertise and influence in multi-user task planning, 2023.

[54] Ruinian Xu, Hongyi Chen, Yunzhi Lin, and Patricio A. Vela. Sgl: Symbolic goal learning in a hybrid, modular framework for human instruction following, 2022.

[55] Minseo Kwon, Yaesol Kim, and Young J. Kim. Fast and accurate task planning using neuro-symbolic language models and multi-level goal decomposition, 2024.

[56] Caelan Reed Garrett, Rohan Chitnis, Rachel Holladay, Beomjoon Kim, Tom Silver, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Integrated task and motion planning. *Annual review of control, robotics, and autonomous systems*, 4(1):265–293, 2021.

[57] Wenhao Yu, Jie Peng, Yueliang Ying, Sai Li, Jianmin Ji, and Yanyong Zhang. Mhrc: Closed-loop decentralized multi-heterogeneous robot collaboration with large language models, 2024.

[58] Yuqian Jiang, Shiqi Zhang, Piyush Khandelwal, and Peter Stone. Task planning in robotics: an empirical comparison of pddl-based and asp-based systems, 2019.

[59] Mert İnan, Aishwarya Padmakumar, Spandana Gella, Patrick Lange, and Dilek Hakkani-Tur. Multimodal contextualized plan prediction for embodied task completion, 2023.

[60] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45, 2020.

[61] Xiaowen Sun, Xufeng Zhao, Jae Hee Lee, Wenhao Lu, Matthias Kerzel, and Stefan Wermter. Details make a difference: Object state-sensitive neurorobotic task planning, 2024.

[62] Gawon Choi and Hyemin Ahn. Can only llms do reasoning?: Potential of small language models in task planning, 2024.

[63] Hongxin Li, Jingran Su, Yuntao Chen, Qing Li, and Zhaoxiang Zhang. Sheetcopilot: Bringing software productivity to the next level through large language models, 2023.

[64] Po-Lin Chen and Cheng-Shang Chang. Interact: Exploring the potentials of chatgpt as a cooperative agent, 2023.

[65] Yunfan Jiang, Agrim Gupta, Zichen Zhang, Guanzhi Wang, Yongqiang Dou, Yanjun Chen, Li Fei-Fei, Anima Anandkumar, Yuke Zhu, and Linxi Fan. Vima: General robot manipulation with multimodal prompts, 2023.

[66] Michael Hess. Mixed-level knowledge representation and variable-depth inference in natural language processing, 1999.

[67] Zidan Wang, Rui Shen, and Bradly Stadie. Wonderful team: Zero-shot physical task planning with visual llms, 2025.

[68] Muhammad Fadhil Ginting, Dong-Ki Kim, Sung-Kyun Kim, Bandi Jai Krishna, Mykel J. Kochenderfer, Shayegan Omidshafiei, and Ali akbar Agha-mohammadi. Saycomply: Grounding field robotic tasks in operational compliance through retrieval-based language models, 2024.

[69] Yoshiki Obinata, Haoyu Jia, Kento Kawaharazuka, Naoaki Kanazawa, and Kei Okada. Remote life support robot interface system for global task planning and local action expansion using foundation models, 2024.

[70] Yingdong Hu, Fanqi Lin, Tong Zhang, Li Yi, and Yang Gao. Look before you leap: Unveiling the power of gpt-4v in robotic vision-language planning, 2023.

[71] Sharif Ahmed, Arif Ahmed, and Nasir U. Eisty. Automatic transformation of natural to unified modeling language: A systematic review, 2022.

[72] Shengbin Yue, Siyuan Wang, Wei Chen, Xuanjing Huang, and Zhongyu Wei. Synergistic multi-agent framework with trajectory learning for knowledge-intensive tasks, 2025.

[73] Lance Ying, Jason Xinyu Liu, Shivam Aarya, Yizirui Fang, Stefanie Tellex, Joshua B. Tenenbaum, and Tianmin Shu. Siftom: Robust spoken instruction following through theory of mind, 2024.

[74] Jun Wang, Jiaming Tong, Kaiyuan Tan, Yevgeniy Vorobeychik, and Yiannis Kantaros. Conformal temporal logic planning using large language models, 2024.

[75] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*, 2022.

[76] Johannes Bjerva. Multitask and multilingual modelling for lexical analysis, 2018.

[77] Zeyu Gao, Yao Mu, Jinye Qu, Mengkang Hu, Lingyue Guo, Ping Luo, and Yanfeng Lu. Dag-plan: Generating directed acyclic dependency graphs for dual-arm cooperative planning, 2024.

[78] Shashank Shekhar, Anthony Favier, and Rachid Alami. An epistemic human-aware task planner which anticipates human beliefs and decisions, 2024.

[79] Yanming Liu, Xinyue Peng, Jiannan Cao, Shi Bo, Yuwei Zhang, Xuhong Zhang, Sheng Cheng, Xun Wang, Jianwei Yin, and Tianyu Du. Tool-planner: Task planning with clusters across multiple tools, 2025.

[80] Chang Liu, Jessica B. Hamrick, Jaime F. Fisac, Anca D. Dragan, J. Karl Hedrick, S. Shankar Sastry, and Thomas L. Griffiths. Goal inference improves objective and perceived performance in human-robot collaboration, 2018.

[81] Nikhil Mehta, Milagro Teruel, Patricio Figueroa Sanz, Xin Deng, Ahmed Hassan Awadallah, and Julia Kiseleva. Improving grounded language understanding in a collaborative environment by interacting with agents through help feedback, 2024.

[82] Qinhao Zhou, Zihan Zhang, Xiang Xiang, Ke Wang, Yuchuan Wu, and Yongbin Li. Enhancing the general agent capabilities of low-parameter llms through tuning and multi-branch reasoning, 2024.

[83] Francisco Martín, Jonatan Ginés, Vicente Matellán, and Francisco J. Rodríguez. Plansys2: A planning system framework for ros2, 2021.

[84] Jun Wang, Guocheng He, and Yiannis Kantaros. Probabilistically correct language-based multi-robot planning using conformal prediction, 2024.

[85] Venkata Naren Devarakonda, Ali Umut Kaypak, Shuaihang Yuan, Prashanth Krishnamurthy, Yi Fang, and Farshad Khorrami. Multitalk: Introspective and extrospective dialogue for human-environment-llm alignment, 2024.

[86] Ruohan Zhang, Faraz Torabi, Garrett Warnell, and Peter Stone. Recent advances in leveraging human guidance for sequential decision-making tasks, 2021.

[87] Chuanneng Sun, Songjun Huang, and Dario Pompili. Retrieval-augmented hierarchical in-context reinforcement learning and hindsight modular reflections for task planning with llms, 2024.

[88] Eric Hsiung, Hiloni Mehta, Junchi Chu, Xinyu Liu, Roma Patel, Stefanie Tellex, and George Konidaris. Generalizing to new domains by mapping natural language to lifted ltl, 2022.

[89] Chen Wang, Danfei Xu, and Li Fei-Fei. Generalizable task planning through representation pretraining, 2022.

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.