On the Prospects of Incorporating Large Language Models (LLMs) in Automated Planning and Scheduling (APS)

Vishal Pallagani[1], Bharath Chandra Muppasani[1], Kaushik Roy[1], Francesco Fabiano[2], Andrea Loreggia[3], Keerthiram Murugesan[4], Biplav Srivastava[1], Francesca Rossi[4], Lior Horesh[4], Amit Sheth[1]

[1]University of South Carolina
[2]New Mexico State University
[3]University of Brescia
[4]IBM Research
{vishalp, bharath, kaushikr}@email.sc.edu, ffabiano@nmsu.edu, andrea.loreggia@gmail.com, keerthiram.murugesan@ibm.com, biplav.s@sc.edu, francesca.rossi2@ibm.com, lhoresh@us.ibm.com

Abstract

Automated Planning and Scheduling is among the growing areas in Artificial Intelligence (AI) where mention of LLMs has gained popularity. Based on a comprehensive review of 126 papers, this paper investigates eight categories based on the unique applications of LLMs in addressing various aspects of planning problems: language translation, plan generation, model construction, multi-agent planning, interactive planning, heuristics optimization, tool integration, and brain-inspired planning. For each category, we articulate the issues considered and existing gaps. A critical insight resulting from our review is that the true potential of LLMs unfolds when they are integrated with traditional symbolic planners, pointing towards a promising neuro-symbolic approach. This approach effectively combines the generative aspects of LLMs with the precision of classical planning methods. By synthesizing insights from existing literature, we underline the potential of this integration to address complex planning challenges. We aim to keep the categorization of papers updated on https://ai4society.github.io/LLM-Planning-Viz/, a collaborative resource that allows researchers to contribute and add new literature to the categorization.

Introduction

As a sub-field of Artificial Intelligence (Russell and Norvig 2003), Automated Planning and Scheduling (Ghallab, Nau, and Traverso 2004) refers to developing algorithms and systems to generate plans or sequences of actions to achieve specific goals in a given environment or problem domain. APS is a valuable tool in domains where there is a need for intelligent decision-making, goal achievement, and efficient resource utilization. It enables the automation of complex tasks, making systems more capable and adaptable in dynamic environments. Over time, APS has evolved from the early development of robust theoretical foundations to practical applications in diverse sectors like manufacturing, space exploration, and personal scheduling.

In parallel with advancements in APS, the development and proliferation of LLMs have marked a substantial leap in AI, particularly within computational linguistics. Evolving from early efforts in natural language processing (NLP), LLMs have undergone significant transformation. Initially focused on basic tasks like word prediction and syntax analysis, newer models are characterized by their ability to generate coherent, contextually relevant text and perform diverse, complex linguistic tasks. Trained on extensive text corpora, LLMs have mastered human-like language patterns. Their recent success in various NLP tasks has prompted efforts to apply these models in APS. There is a notable shift towards using language constructs to specify aspects of planning, such as preconditions, effects, and goals, rather than relying solely on traditional planning domain languages like PDDL.

This paper presents an exhaustive literature review exploring the integration of LLMs in APS across eight categories: Language Translation, Plan Generation, Model Construction, Multi-agent Planning, Interactive Planning, Heuristics Optimization, Brain-Inspired Planning, and Tool Integration. Table 1 describes the eight categories. Our comprehensive analysis of 126 papers categorizes LLMs' diverse contributions and identifies significant gaps in each domain. Through our review, we put forward the following position:

Position Statement

Integrating LLMs into APS marks a pivotal advancement, bridging the gap between the advanced reasoning of traditional APS and the nuanced language understanding of LLMs. Traditional APS systems excel in structured, logical planning but often lack flexibility and contextual adaptability, a gap readily filled by LLMs. Conversely, while LLMs offer unparalleled natural language processing and a vast knowledge base, they fail to generate precise, actionable plans where APS systems thrive. This integration surpasses the limitations of each standalone method, offering a dynamic and context-aware planning approach while also scaling up the traditional use of data and past experiences in the planning process.

In the forthcoming sections, we delve into the background

Category	Description
Language Translation	Involves converting natural language into structured planning languages or formats like PDDL and vice-versa, enhancing the interface between human linguistic input and machine-understandable planning directives.
Plan Generation	Entails the creation of plans or strategies directly by LLMs, focusing on generating actionable sequences or decision-making processes.
Model Construction	Utilizes LLMs to construct or refine world and domain models essential for accurate and effective planning.
Multi-agent Planning	Focuses on scenarios involving multiple agents, where LLMs contribute to coordination and cooperative strategy development.
Interactive Planning	Centers on scenarios requiring iterative feedback or interactive planning with users, external verifiers, or environment, emphasizing the adaptability of LLMs to dynamic inputs.
Heuristics Optimization	Applies LLMs in optimizing planning processes through refining existing plans or providing heuristic assistance to symbolic planners.
Tool Integration	Encompasses studies where LLMs act as central orchestrators or coordinators in a tool ecosystem, interfacing with planners, theorem provers, and other systems.
Brain-Inspired Planning	Covers research focusing on LLM architectures inspired by neurological or cognitive processes, particularly to enhance planning capabilities.

Table 1: Comprehensive description of the eight categories utilizing LLMs in APS

of LLMs and classical planning problems, accompanied by the identification of literature. This sets the stage for an in-depth exploration of the application of LLMs in APS, where we critically examine the strengths and limitations of LLMs. Our position on the emerging neuro-symbolic AI paradigm is central to our discussion, highlighting its unique advantages over purely neural network-based (i.e., statistical AI) or symbolic AI approaches. Finally, we will discuss prospective developments, address potential challenges, and identify promising opportunities in the field.

Background

Large Language Models

Large language models are neural network models with upwards of ~ 3 billion parameters that are trained on extremely large corpora of natural language data (trillions of tokens/-words). These models are proficient in interpreting, generating, and contextualizing human language, leading to applications ranging from text generation to language-driven reasoning tasks. The evolution of LLMs in NLP began with rule-based models, progressed through statistical models, and achieved a significant breakthrough with the introduction of neural network-based models. The shift to sequence-based neural networks, with Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, marked a notable advancement due to their capability to process information and context over long sequences. Shortcomings in RNNs and LSTMs due to vanishing gradients and, consequently, loss of *very long* sequence contexts lead to the transformer model, which introduced self-attention (SA) mechanisms. The SA mechanism enabled focus on different parts of a long input sequence in parallel, which enhanced understanding of contextual nuances in language patterns over extremely long sequences. The SA mechanism is also complemented with positional encodings in transformers to enable the model to maintain an awareness of word/token order, which is required to understand accurate grammar and syntax. The self-attention mechanism, central to transformers, uses a query, key, and value system to contextualize dependencies in the input sequence. Informally, the SA concept is inspired by classical information retrieval systems where the query is the input sequence context, the key refers to a "database" contained within the parametric memory, and the value is the actual value present at that reference. The operation is mathematically expressed in Equation 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

In this equation, Q, K, and V denote the query, key, and value matrices. The scaling factor $\sqrt{d_k}$, where d_k is the dimension of the keys, is employed to standardize the vectors to unit variance for ensuring stable softmax gradients during training. Since the introduction of LLMs with self-attention, there have been several architectural variants depending on the downstream tasks.

Causal Language Modeling (CLMs): CLMs, such as GPT-4, are designed for tasks where text generation is sequential and dependent on the preceding context. They predict each subsequent word based on the preceding words, modeling the probability of a word sequence in a forward direction. This process is mathematically formulated as shown in Equation 2.

$$P(T) = \prod_{i=1}^{n} P(t_i|t_{<i}) \tag{2}$$

In this formulation, $P(t_i|t_{<i})$ represents the probability of the i-th token given all preceding tokens, $t_{<i}$. This characteristic makes CLMs particularly suitable for applications like content generation, where the flow and coherence of the text in the forward direction are crucial.

Masked Language Modeling (MLMs): Unlike CLMs, MLMs like BERT are trained to understand the bidirectional context by predicting words randomly masked in a sentence. This approach allows the model to learn both forward and backward dependencies in language structure. The MLM prediction process can be represented as Equation 3.

$$P(T_{\text{masked}}|T_{\text{context}}) = \prod_{i \in M} P(t_i|T_{\text{context}}) \tag{3}$$

Here, T_{masked} is the set of masked tokens in the sentence, T_{context} represents the unmasked part of the sentence, and M is the set of masked positions. MLMs have proven effective in NLP tasks such as sentiment analysis or question answering.

Sequence-to-Sequence (Seq2Seq) Modeling: Seq2Seq models, like T5, are designed to transform an input sequence into a related output sequence. They are often employed in tasks that require a mapping between different types of sequences, such as language translation or summarization. The Seq2Seq process is formulated as Equation 4.

$$P(T_{\text{output}}|T_{\text{input}}) = \prod_{i=1}^{m} P(t_{\text{output}_i}|T_{\text{input}}, t_{\text{output}_{<i}}) \tag{4}$$

In Equation 4, T_{input} is the input sequence, T_{output} is the output sequence, and $P(t_{\text{output}_i}|T_{\text{input}}, t_{\text{output}_{<i}})$ calculates the probability of generating each token in the output sequence, considering both the input sequence and the preceding tokens in the output sequence.

In addition to their architectural variants, the utility of LLMs is further enhanced by specific model utilization strategies, enabling their effective adaptation to various domains at scale. One key strategy is fine-tuning, which applies to pre-trained LLMs. Pre-trained LLMs are models already trained on large datasets to understand and generate language, acquiring a broad linguistic knowledge base. Fine-tuning involves further training pre-trained LLMs on a smaller, task-specific dataset, thereby adjusting the neural network weights for particular applications. This process is mathematically represented in Equation 5.

$$\theta_{\text{fine-tuned}} = \theta_{\text{pre-trained}} - \eta \cdot \nabla_\theta L(\theta, D_{\text{task}}) \tag{5}$$

Here, $\theta_{\text{fine-tuned}}$ are the model parameters after fine-tuning, $\theta_{\text{pre-trained}}$ are the parameters obtained from pre-training, η is the learning rate, and $\nabla_\theta L(\theta, D_{\text{task}})$ denotes the gradient of the loss function L with respect to the parameters θ on the task-specific dataset D_{task}.

$$P(T|C) = \prod_{i=1}^{n} P(t_i|t_{<i}, C) \tag{6}$$

Complementing the fine-tuning approach is in-context learning, an alternative strategy that is particularly characteristic of models like the GPT series. This method diverges from fine-tuning by enabling the model to adapt its responses based on immediate context or prompts without necessitating further training. The efficacy of in-context learning is a direct consequence of the comprehensive pre-training phase, where models are exposed to diverse textual datasets, thereby acquiring a nuanced understanding of language and context. Given a context C, the model generates text T that is contextually relevant, as shown in Equation 6. Here, $P(T|C)$ is the probability of generating text T given the context C, and $P(t_i|t_{<i}, C)$ is the probability of generating the i-th token t_i given the preceding tokens $t_{<i}$ and the context C.

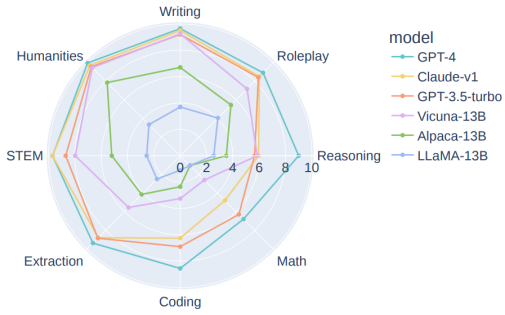

Figure 1: Radar chart showcasing the relative performance of six language models (GPT-4, Claude-v1, GPT-3.5-turbo, Vicuna-13B, Alpaca-13B, LLama-13B) across key domains: Writing, Roleplay, Reasoning, Math, Coding, Extraction, STEM, and Humanities from Zheng et al. (2023a).

These diverse model types and training methodologies under the umbrella of LLMs showcase the flexibility and adaptability of language models in handling a wide range of complex tasks. Figure 1 illustrates the comparative capabilities of different LLMs across various competency domains, such as Writing (evaluating text generation quality), Roleplay (assessing conversational interaction), Reasoning (logical problem-solving), Math (numerical problem-solving), Coding (programming language understanding and generation), Extraction (information retrieval from text), STEM (proficiency in scientific and technical contexts), and Humanities (engagement with arts, history, and social sciences content). Across these domains, GPT-4 exhibits the strongest performance in the benchmark dataset evaluated by Zheng et al. (2023a), indicative of its superior training and extensive knowledge base. Expanding LLMs into applications such as code generation signifies their adaptability and potential for cross-disciplinary innovation. However, fine-tuning and in-context learning methodologies also bring challenges, such as potential data overfitting and reliance on the quality of input context. LLMs' continuous develop-

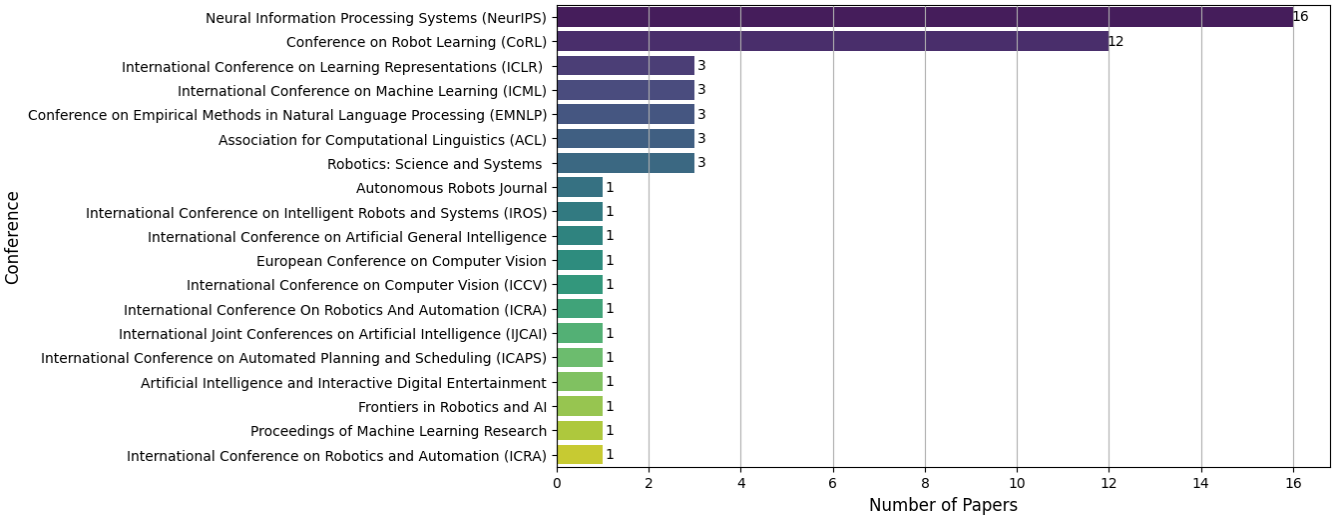

Figure 2: Of the 126 papers surveyed in this study, 55 were accepted by peer-reviewed conferences. This chart illustrates the distribution of these papers across various conferences in the fields of LLMs and APS, highlighting the primary forums for scholarly contributions in these areas.

ment and refinement promise to open new frontiers in various domains, including automated planning and scheduling, by bridging AI with human-like language understanding.

Automated Planning and Scheduling

APS is a branch of AI that focuses on the creation of strategies or action sequences, typically for execution by intelligent agents, autonomous robots, and unmanned vehicles. A basic category in APS is a Classical Planning Problem (CPP) (Russell and Norvig 2003) which is a tuple $\mathcal{M} = \langle \mathcal{D}, \mathcal{I}, \mathcal{G} \rangle$ with domain $\mathcal{D} = \langle F, A \rangle$ - where F is a set of fluents that define a state $s \subseteq F$, and A is a set of actions - and initial and goal states $\mathcal{I}, \mathcal{G} \subseteq F$. Action $a \in A$ is a tuple $(c_a, \mathit{pre}(a), \mathit{eff}^{\pm}(a))$ where c_a is the cost, and $\mathit{pre}(a), \mathit{eff}^{\pm}(a) \subseteq F$ are the preconditions and add/delete effects, i.e., $\delta_{\mathcal{M}}(s, a) \models \perp s$ *if* $s \not\models \mathit{pre}(a)$; *else* $\delta_{\mathcal{M}}(s, a) \models s \cup \text{eff}^{+}(a) \setminus \text{eff}^{-}(a)$ where $\delta_{\mathcal{M}}(\cdot)$ is the transition function. The cumulative transition function is $\delta_{\mathcal{M}}(s, (a_1, a_2, \ldots, a_n)) = \delta_{\mathcal{M}}(\delta_{\mathcal{M}}(s, a_1), (a_2, \ldots, a_n))$. A plan for a CPP is a sequence of actions $\langle a_1, a_2, \ldots, a_n \rangle$ that transforms the initial state $\mathcal{I}$ into the goal state $\mathcal{G}$ using the transition function $\delta_{\mathcal{M}}$. Traditionally, a CPP is encoded using a symbolic representation, where states, actions, and transitions are explicitly enumerated. This symbolic approach, often implemented using Planning Domain Definition Language or PDDL (McDermott et al. 1998), ensures precise and unambiguous descriptions of planning problems. This formalism allows for applying search algorithms and heuristic methods to find a sequence of actions that lead to the goal state, which is the essence of the plan.

The advent of LLMs has sparked a significant evolution in representation methods for CPPs, moving towards leveraging the expressive power of natural language (Valmeekam et al. 2023a) and the perceptual capabilities of vision (Asai 2018). These novel approaches, inherently more suited for LLM processing, use text and vision-based representations, allowing researchers to utilize the pre-existing knowledge within LLMs. This shift enables a more humanistic comprehension and reasoning about planning tasks, enhancing the flexibility and applicability of planning algorithms in complex, dynamic environments. LLMs, while distinct in being trained on vast datasets outside the traditional scope of planning, loosely connect to previous data-driven methodologies, such as case-based reasoning (Xu 1995) applied to planning and Hierarchical Task Network (HTN) (**?**) which make use of task knowledge. It is an open area of how LLMs may be used synergistically with prior methods.

LLMs in APS – Literature Selection

A comprehensive survey of existing literature was conducted to explore the application of LLMs for automated planning. This endeavor identified 126 pertinent research papers showcasing various methodologies, applications, and theoretical insights into utilizing LLMs within this domain.

The selection of these papers was guided by stringent criteria, focusing primarily on their relevance to the core theme of LLMs in automated planning. The search, conducted across multiple academic databases and journals, was steered by keywords such as “Large Language Models”, “Automated Planning”, “LLMs in Planning”, and “LLMs + Robotics”. Figure 2 presents the distribution of these selected papers across various peer-reviewed conferences, underlining the breadth and diversity of forums addressing the intersection of LLMs and APS. Even if a paper originated from a workshop within a conference, only the conference name is listed. Out of 126 papers, 71 are under review or available on arXiv. The inclusion criteria prioritized the relevance and contribution of papers to automated planning with LLMs over the publication date. Nonetheless, all surveyed papers emerged from either 2022 or 2023, with 12

papers from 2022 and 114 papers from 2023, underscoring the recent surge in LLM research. A word cloud was generated to visually capture the prevalent research themes reflected in these papers' titles, illustrated in Figure 3. This cloud highlights the frequent use of terms such as "Language Model" and "Planning", which dominate the current discourse. In contrast, the emergence of "Neuro-Symbolic" reflects a nascent yet growing interest in integrating neural and symbolic approaches within the field. This systematic approach ensured a comprehensive inclusion of seminal works and recent advancements.

Upon the accumulation of these papers, a meticulous manual categorization was undertaken. The papers were divided into four piles, each containing approximately 30 papers. Each pile was manually categorized by one author, with the final categorization being reviewed by all authors. Each paper could belong to multiple categories out of the eight established during this process. The maximum number of categories assigned to a single paper was three, although the median was typically one category per paper. This process was pivotal in distilling the vast information into coherent, thematic groups. The categorization was conducted based on the specific application of LLMs in planning. This formed eight distinct categories, each representing a unique facet of LLM application in automated planning. These categories facilitate a structured analysis and highlight LLMs' diverse applications and theoretical underpinnings in this field.

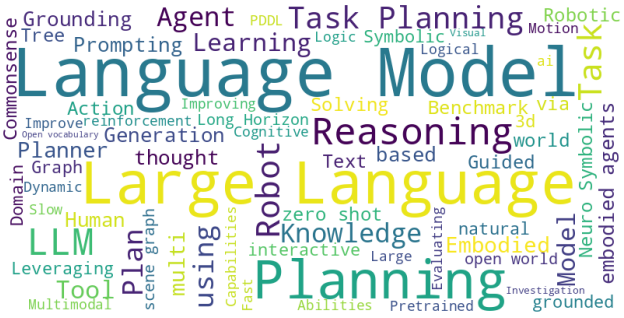

Figure 3: Word cloud of terms from the titles of papers surveyed in this study, displaying the prevalence of "Language Model" and "Planning" as central themes. The presence of "Neuro-Symbolic" indicates an emergent trend toward the fusion of neural and symbolic methodologies in the domain.

LLMs in APS – Literature Discussion

This section dwelves into the diverse applications of LLMs in planning tasks. We have identified eight distinct categories based on the utility and application of LLMs in planning, which are concisely summarized in Table 1. Figure 4 provides a detailed taxonomy, illustrating the categorization of the identified research papers.

Language Translation

Language translation in the context of LLMs and planning involves transforming natural language instructions into structured planning languages (Wong et al. 2023; Kelly et al. 2023; Yang 2023; Pan et al. 2023; Xie et al. 2023; Yang, Ishay, and Lee 2023; Lin et al. 2023c; Sakib and Sun 2023; Yang et al. 2023b; Parakh et al. 2023; Yang et al. 2023a; Dai et al. 2023; Ding et al. 2023b; Zelikman et al. 2023; Xu et al. 2023b; Chen et al. 2023a; You et al. 2023) such as PDDL, and vice versa, utilizing in-context learning techniques (Guan et al. 2023). This capability effectively bridges the gap between human linguistic expression and machine-understandable formats, enhancing intuitive and efficient planning processes. The LLM+P framework (Liu et al. 2023) exemplifies this by converting natural language descriptions of planning problems into PDDL using GPT-4, leveraging classical planners for solution finding, and then translating these solutions back into natural language, with a specific focus on robot planning scenarios. Additionally, Graph2NL (Chalvatzaki et al. 2023) generates natural language text from scene graphs for long-horizon robot reasoning tasks, while (Shirai et al. 2023) introduces a vision-to-language interpreter for robot task planning. Further, (Brohan et al. 2023) examines the grounding of LLM-generated natural language utterances in actionable robot tasks, and (Yang, Gaglione, and Topcu 2022) utilizes LLMs for creating finite-state automatons for sequential decision-making problems. Despite these advancements, a critical research gap emerges in the autonomous translation capabilities of LLMs, particularly in converting natural language to PDDL without external expert intervention.

> While LLMs effectively translate PDDL to natural language, **a notable gap is evident in their limited understanding of real-world objects and the problem of grounding affordances**, mainly when translating natural language to structured languages like PDDL. Addressing this gap calls for integrating neuro-symbolic approaches in LLMs, where the fusion of perceptual experience for concrete concept understanding from knowledge graphs complements LLMs' proficiency in distributional statistics (Lenat and Marcus 2023).

Plan Generation

This category focuses on directly generating plans using LLMs. The research, primarily utilizing causal language models through in-context learning (Sermanet et al. 2023; Li et al. 2023b; Silver et al. 2023; Parakh et al. 2023; Zelikman et al. 2023; Besta et al. 2023; Huang et al. 2023a; Dalal et al. 2023; Wang et al. 2023b; Valmeekam et al. 2022; Valmeekam, Marquez, and Kambhampati 2023; Gramopadhye and Szafir 2022; Singh et al. 2023), demonstrates modest performance, indicating notable challenges in employing LLMs for effective plan generation. Novel in-context learning strategies, such as the Chain-of-Symbol and Tree of Thoughts, have been introduced to enhance LLMs' reasoning capabilities (Hu et al. 2023b; Yao et al. 2023). Efforts to generate multimodal, text, and image-based goal-conditioned plans are exemplified by (Lu et al. 2023b). Additionally, a subset of studies in this survey investigates the fine-tuning of seq2seq, code-based language models (Pallagani et al. 2022, 2023b), which are noted for their advanced

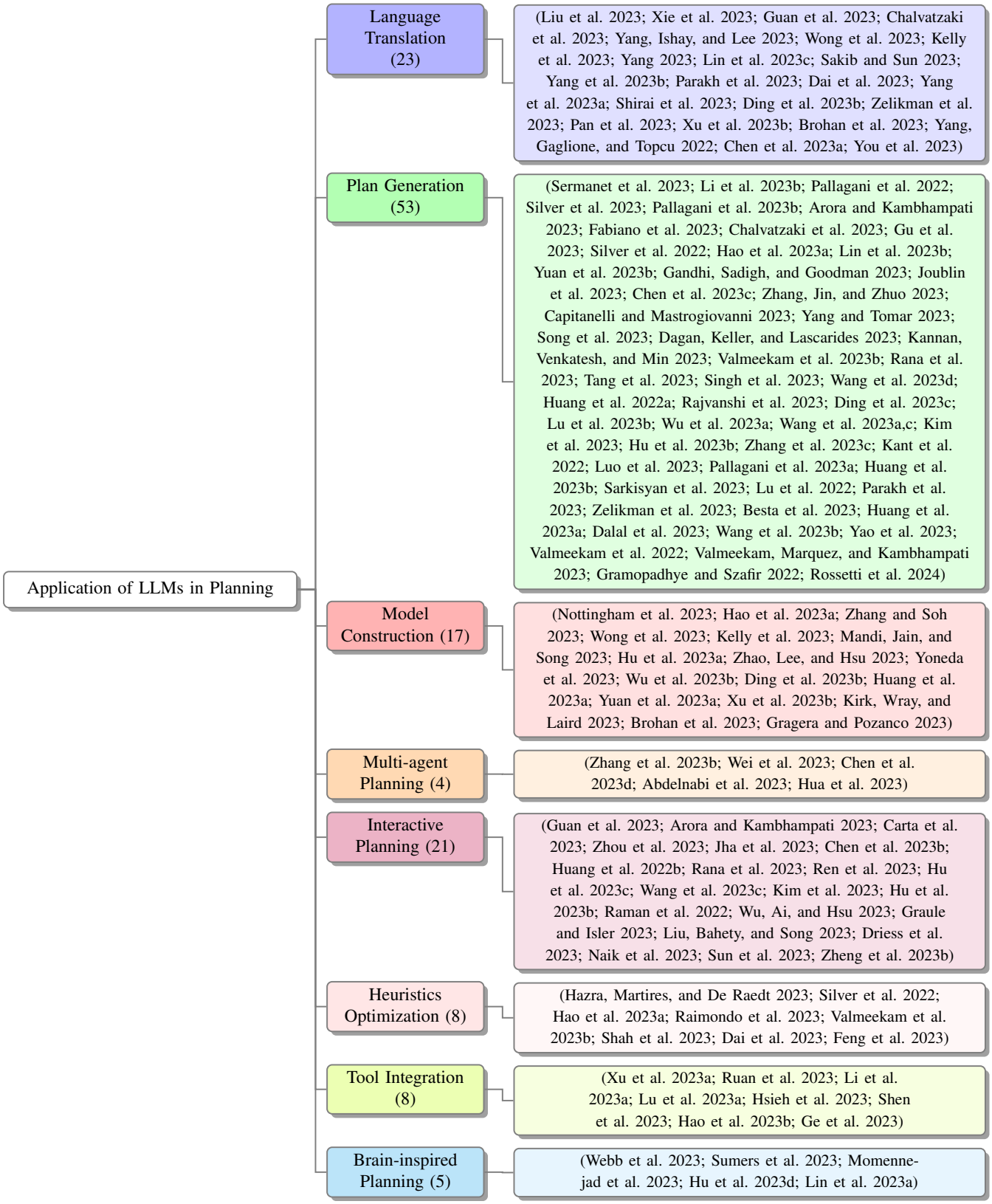

Figure 4: Taxonomy of recent research in the intersection of LLMs and Planning into categories (#). Each has scholarly papers based on their unique application or customization of LLMs in addressing various aspects of planning problems.

syntactic encoding. These models show promise in improving plan generation within the confines of their training datasets (Logeswaran et al. 2023), yet exhibit limitations in generalizing to out-of-distribution domains (Pallagani et al. 2023a), which is addressed in (Rossetti et al. 2024).

Causal LLMs are predominantly used for plan generation, but their performance is often **limited due to their design, which is focused on generating text based on preceding input.** On the other hand, seq2seq LLMs can generate valid plans but **struggle with generalization across diverse domains.** This limitation highlights an opportunity for a synergistic approach: integrating even imperfect LLM outputs with symbolic planners can expedite heuristic searches, thereby enhancing efficiency and reducing search times (Fabiano et al. 2023).

Model Construction

This category employs LLMs to build or refine world and domain models essential for accurate planning. Nottingham et al. (2023); Yuan et al. (2023a) leverage in-context learning with LLMs to develop an abstract world model in the Minecraft domain, highlighting the challenge of semantic grounding in LLMs. Similarly, Gragera and Pozanco (2023) explore the capability of LLMs in completing ill-defined PDDL domains. Efforts such as (Huang et al. 2023a; Brohan et al. 2023) delve into LLMs' grounding capabilities, with SayCan (Brohan et al. 2023) notably achieving 74% executable plans. Hao et al. (2023a); Yoneda et al. (2023) innovatively positions LLMs as both world models and reasoning agents, enabling the simulation of world states and prediction of action outcomes. Research by (Zhang and Soh 2023; Wong et al. 2023; Mandi, Jain, and Song 2023; Hu et al. 2023a; Zhao, Lee, and Hsu 2023; Ding et al. 2023b; Huang et al. 2023a; Wu et al. 2023b; Xu et al. 2023b; Brohan et al. 2023) shows that LLMs can effectively model high-level human states and behaviors using their commonsense knowledge. Yet, they face difficulties accurately processing low-level geometrical or shape features due to spatial and numerical reasoning constraints. Additionally, Kelly et al. (2023) investigates the potential of LLMs in conjunction with planners to craft narratives and logical story models, integrating human-in-the-loop for iterative edits.

LLMs often **struggle with detailed spatial reasoning and processing low-level environmental features, limiting their effectiveness in model construction.** Integrating world models presents a viable solution, offering advanced abstractions for reasoning that encompass human-like cognitive elements and interactions, thereby enhancing LLMs' capabilities in model construction (Hu and Shu 2023).

Multi-agent Planning

In multi-agent planning, LLMs play a vital role in scenarios involving interaction among multiple agents, typically modeled using distinct LLMs. These models enhance coordination and cooperation, leading to more complex and effective multi-agent strategies. (Zhang et al. 2023b) introduces an innovative framework that employs LLMs to develop cooperative embodied agents. AutoGraph (Wei et al. 2023) leverages LLMs to generate autonomous agents adept at devising solutions for varied graph-structured data problems. Addressing scalability in multi-robot task planning, (Chen et al. 2023d) proposes frameworks for the collaborative function of different LLM-based agents. Furthermore, (Abdelnabi et al. 2023) and (Hua et al. 2023) collectively demonstrate the effectiveness of LLM agents in complex negotiation and decision-making environments.

A key gap in multi-agent planning with LLMs lies in **standardizing inter-agent communication and maintaining distinct belief states, including human aspects.** Overcoming this requires advanced LLM algorithms for dynamic alignment of communication and belief states, drawing on epistemic reasoning and knowledge representation (de Zarzà et al. 2023).

Interactive Planning

In this category, LLMs are utilized in dynamic scenarios where real-time adaptability to user feedback or iterative planning is essential. The refinement of LLM outputs is typically achieved through four primary feedback variants: **(a)** External verifiers, such as VAL(Howey, Long, and Fox 2004) for PDDL or scene descriptors and success detectors in robotics (Guan et al. 2023; Arora and Kambhampati 2023; Jha et al. 2023; Huang et al. 2022b; Liu, Bahety, and Song 2023; Rana et al. 2023; Ren et al. 2023; Kim et al. 2023; Graule and Isler 2023; Driess et al. 2023; Zheng et al. 2023b); **(b)** Online reinforcement learning, which progressively updates the LLM about environmental changes (Carta et al. 2023); **(c)** Self-refinement by LLMs, where they provide feedback on their own outputs (Zhou et al. 2023; Hu et al. 2023c,b; Ding et al. 2023a; Sun et al. 2023; Naik et al. 2023); **(d)** Input from human experts (Raman et al. 2022; Wu, Ai, and Hsu 2023). Furthermore, (Chen et al. 2023b) introduces the "Action Before Action" method, enabling LLMs to proactively seek relevant information from external sources in natural language, thereby improving embodied decision-making in LLMs by 40%.

A key gap in interactive planning with LLMs lies in harmonizing the "fast" neural processing of LLMs with "slow" symbolic reasoning, as manifested in feedback mechanisms. This integration is key to **maintaining the neural speed of LLMs while effectively embedding the depth and precision of feedback**, which is vital for accuracy in dynamic planning scenarios (Zhang et al. 2023a).

Heuristics Optimization

In Heuristics Optimization, LLMs are leveraged to enhance planning processes, either by refining existing plans or aiding symbolic planners with heuristic guidance. Studies like (Hazra, Martires, and De Raedt 2023; Hao et al. 2023a; Dai et al. 2023; Feng et al. 2023) have effectively coupled LLMs with heuristic searches to identify optimal action sequences. Research by (Silver et al. 2022; Shah et al. 2023; Valmeekam

et al. 2023b) reveals that LLMs' outputs, even if partially correct, can provide valuable direction for symbolic planners such as LPG (Gerevini and Serina 2002), especially in problems beyond the LLMs' solvable scope. Furthermore, (Raimondo et al. 2023) makes an intriguing observation that including workflows and action plans in LLM input prompts can notably enhance task-oriented dialogue generalization.

This category marks significant progress towards realizing neuro-symbolic approaches in APS. **Current methods emphasize plan validity, often at the expense of time efficiency.** Future research should look at how to continually evolve LLMs for better plan generation, with its experience from complimenting symbolic planners (Du et al. 2023).

Tool Integration

In tool integration, LLMs are coordinators within various planning tools, enhancing functionality in complex scenarios. Studies like (Xu et al. 2023a; Lu et al. 2023a; Shen et al. 2023; Hao et al. 2023b; Ge et al. 2023) demonstrate that incorporating tools such as web search engines, Python functions, and API endpoints enhances LLM reasoning abilities. However, (Ruan et al. 2023) notes LLMs tend to overrely on specific tools, potentially prolonging the planning process. (Li et al. 2023a) introduces a benchmark for tool-augmented LLMs. While typical approaches involve teaching LLMs tool usage via multiple prompts, (Hsieh et al. 2023) suggests that leveraging tool documentation offers improved planning capabilities, circumventing the need for extensive demonstrations.

LLMs often **hallucinate non-existent tools, overuse a single tool, and face scaling challenges with multiple tools.** Overcoming these issues is key to enabling LLMs to select and utilize various tools in complex planning scenarios effectively (Elaraby et al. 2023).

Brain-Inspired Planning

This area explores neurologically and cognitively inspired architectures in LLMs (Webb et al. 2023; Sumers et al. 2023; Momennejad et al. 2023; Hu et al. 2023d; Lin et al. 2023a), aiming to replicate human-like planning in enhancing problem-solving. However, while these methods rely on in-context learning, they frequently encounter challenges such as hallucination and grounding, as previously discussed, and tend to be more computationally intensive than in-context learning alone.

While LLMs attempt to mimic symbolic solvers through in-context learning for brain-inspired modules, this approach **lacks adaptability and is a superficial understanding of complex cognitive processes.** To overcome these issues, developing systems where neural and symbolic components are intrinsically intertwined is critical as it would accurately mirror human cognitive capabilities in planning (Fabiano et al. 2023).

Discussion and Conclusion

In this position paper, we comprehensively investigate the role of LLMs within the domain of APS, analyzing 126 scholarly articles across eight distinct categories. This extensive survey not only provides a detailed landscape of current LLM applications and their limitations but also highlights the volume of research in each category: Language Translation with 23 papers demonstrates LLMs' proficiency, whereas Plan Generation, the most researched category with 53 papers, reveals their shortcomings in optimality, completeness, and correctness compared to traditional combinatorial planners. Our exploration extends to Model Construction (17 papers), which examines LLMs in developing planning models, and the relatively unexplored area of Multi-agent Planning (4 papers). Interactive Planning is well represented with 21 papers, illustrating LLMs' adaptability in feedback-centric scenarios. Despite being less researched, Heuristics Optimization and Tool Integration, each with 8 papers, provide valuable insights into efficiency enhancement and integration of LLMs with symbolic solvers. Lastly, Brain-inspired Planning, although least represented with 5 papers, opens innovative avenues for human-like planning processes in LLMs. By identifying the research distribution and gaps in these categories, our paper proposes how neuro-symbolic approaches can address these voids, thereby underscoring the varying degrees of LLM applications in APS and guiding future research towards enhancing their capabilities in complex planning tasks.

It is important to acknowledge that while LLMs have shown promise, they are not a panacea for the inherent complexities of automated planning. The expectation that LLMs, operating within polynomial run-time bounds, could supplant the nuanced and often non-polynomial complexities of symbolic planners is not yet realizable. Indeed, the strengths of LLMs do not currently include generating sequences of actions akin to those devised by symbolic planners, which are essential for creating a coherent and practical plan for complex problems. However, this does not diminish the potential utility of LLMs within this space. When considering average-case scenarios, which are typically less complex than worst-case scenarios, LLMs could offer substantial efficiencies. They can be seen as akin to meta-heuristic approaches, capable of accelerating plan generation in a variety of settings. As such, their application, governed by cognitive-inspired frameworks like SOFAI(Fabiano et al. 2023), could delineate when and where their use is most advantageous.

Future research should prioritize three areas: developing new LLM training paradigms that ensure coherence and goal alignment in outputs; delving into Henry Kautz's neuro-symbolic taxonomies (Kautz 2022) to better integrate neural and symbolic methods; and establishing clear performance metrics for LLM-assisted planners. In conclusion, integrating LLMs into automated planning, while challenging, opens avenues for innovation. Embracing a symbiotic approach that combines the creative strengths of LLMs with the precision of symbolic planners can lead to more effective, sophisticated AI applications in planning.

References

Abdelnabi, S.; Gomaa, A.; Sivaprasad, S.; Schönherr, L.; and Fritz, M. 2023. Llm-deliberation: Evaluating llms with interactive multi-agent negotiation games. *arXiv preprint arXiv:2309.17234*.

Arora, D.; and Kambhampati, S. 2023. Learning and Leveraging Verifiers to Improve Planning Capabilities of Pre-trained Language Models. *arXiv preprint arXiv:2305.17077*.

Asai, M. 2018. Photo-Realistic Blocksworld Dataset. *arXiv preprint arXiv:1812.01818*.

Besta, M.; Blach, N.; Kubicek, A.; Gerstenberger, R.; Gianinazzi, L.; Gajda, J.; Lehmann, T.; Podstawski, M.; Niewiadomski, H.; Nyczyk, P.; et al. 2023. Graph of thoughts: Solving elaborate problems with large language models. *arXiv preprint arXiv:2308.09687*.

Brohan, A.; Chebotar, Y.; Finn, C.; Hausman, K.; Herzog, A.; Ho, D.; Ibarz, J.; Irpan, A.; Jang, E.; Julian, R.; et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on Robot Learning*, 287–318. PMLR.

Capitanelli, A.; and Mastrogiovanni, F. 2023. A Framework to Generate Neurosymbolic PDDL-compliant Planners. *arXiv preprint arXiv:2303.00438*.

Carta, T.; Romac, C.; Wolf, T.; Lamprier, S.; Sigaud, O.; and Oudeyer, P.-Y. 2023. Grounding large language models in interactive environments with online reinforcement learning. *arXiv preprint arXiv:2302.02662*.

Chalvatzaki, G.; Younes, A.; Nandha, D.; Le, A. T.; Ribeiro, L. F.; and Gurevych, I. 2023. Learning to reason over scene graphs: a case study of finetuning GPT-2 into a robot language model for grounded task planning. *Frontiers in Robotics and AI*, 10.

Chen, B.; Xia, F.; Ichter, B.; Rao, K.; Gopalakrishnan, K.; Ryoo, M. S.; Stone, A.; and Kappler, D. 2023a. Open-vocabulary queryable scene representations for real world planning. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 11509–11522. IEEE.

Chen, X.; Zhang, S.; Zhang, P.; Zhao, L.; and Chen, J. 2023b. Asking Before Action: Gather Information in Embodied Decision Making with Language Models. *arXiv preprint arXiv:2305.15695*.

Chen, Y.; Arkin, J.; Zhang, Y.; Roy, N.; and Fan, C. 2023c. AutoTAMP: Autoregressive Task and Motion Planning with LLMs as Translators and Checkers. *arXiv preprint arXiv:2306.06531*.

Chen, Y.; Arkin, J.; Zhang, Y.; Roy, N.; and Fan, C. 2023d. Scalable Multi-Robot Collaboration with Large Language Models: Centralized or Decentralized Systems? *arXiv preprint arXiv:2309.15943*.

Dagan, G.; Keller, F.; and Lascarides, A. 2023. Dynamic Planning with a LLM. *arXiv preprint arXiv:2308.06391*.

Dai, Z.; Asgharivaskasi, A.; Duong, T.; Lin, S.; Tzes, M.-E.; Pappas, G.; and Atanasov, N. 2023. Optimal Scene Graph Planning with Large Language Model Guidance. *arXiv preprint arXiv:2309.09182*.

Dalal, M.; Chiruvolu, T.; Chaplot, D. S.; and Salakhutdinov, R. 2023. Plan-Seq-Learn: Language Model Guided RL for Solving Long Horizon Robotics Tasks. In *2nd Workshop on Language and Robot Learning: Language as Grounding*.

de Zarzà, I.; de Curtò, J.; Roig, G.; Manzoni, P.; and Calafate, C. T. 2023. Emergent Cooperation and Strategy Adaptation in Multi-Agent Systems: An Extended Coevolutionary Theory with LLMs. *Electronics*, 12(12): 2722.

Ding, Y.; Zhang, X.; Amiri, S.; Cao, N.; Yang, H.; Kaminski, A.; Esselink, C.; and Zhang, S. 2023a. Integrating Action Knowledge and LLMs for Task Planning and Situation Handling in Open Worlds. *arXiv preprint arXiv:2305.17590*.

Ding, Y.; Zhang, X.; Paxton, C.; and Zhang, S. 2023b. Leveraging Commonsense Knowledge from Large Language Models for Task and Motion Planning. In *RSS 2023 Workshop on Learning for Task and Motion Planning*.

Ding, Y.; Zhang, X.; Paxton, C.; and Zhang, S. 2023c. Task and motion planning with large language models for object rearrangement. *arXiv preprint arXiv:2303.06247*.

Driess, D.; Xia, F.; Sajjadi, M. S.; Lynch, C.; Chowdhery, A.; Ichter, B.; Wahid, A.; Tompson, J.; Vuong, Q.; Yu, T.; et al. 2023. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*.

Du, M.; Luu, A. T.; Ji, B.; and Ng, S.-k. 2023. From Static to Dynamic: A Continual Learning Framework for Large Language Models. *arXiv preprint arXiv:2310.14248*.

Elaraby, M.; Lu, M.; Dunn, J.; Zhang, X.; Wang, Y.; and Liu, S. 2023. Halo: Estimation and reduction of hallucinations in open-source weak large language models. *arXiv preprint arXiv:2308.11764*.

Fabiano, F.; Pallagani, V.; Ganapini, M. B.; Horesh, L.; Loreggia, A.; Murugesan, K.; Rossi, F.; and Srivastava, B. 2023. Fast and Slow Planning. *arXiv preprint arXiv:2303.04283*.

Feng, X.; Wan, Z.; Wen, M.; Wen, Y.; Zhang, W.; and Wang, J. 2023. Alphazero-like tree-search can guide large language model decoding and training. *arXiv preprint arXiv:2309.17179*.

Gandhi, K.; Sadigh, D.; and Goodman, N. D. 2023. Strategic Reasoning with Language Models. *arXiv preprint arXiv:2305.19165*.

Ge, Y.; Hua, W.; Ji, J.; Tan, J.; Xu, S.; and Zhang, Y. 2023. Openagi: When llm meets domain experts. *arXiv preprint arXiv:2304.04370*.

Gerevini, A.; and Serina, I. 2002. LPG: A Planner Based on Local Search for Planning Graphs with Action Costs. In *Aips*, volume 2, 281–290.

Ghallab, M.; Nau, D.; and Traverso, P. 2004. *Automated Planning: Theory and Practice*. The Morgan Kaufmann Series in Artificial Intelligence. Amsterdam: Morgan Kaufmann. ISBN 978-1-55860-856-6.

Gragera, A.; and Pozanco, A. 2023. Exploring the Limitations of using Large Language Models to Fix Planning Tasks.

Gramopadhye, M.; and Szafir, D. 2022. Generating executable action plans with environmentally-aware language models. *arXiv preprint arXiv:2210.04964*.

Graule, M. A.; and Isler, V. 2023. GG-LLM: Geometrically Grounding Large Language Models for Zero-shot Human Activity Forecasting in Human-Aware Task Planning. *arXiv preprint arXiv:2310.20034*.

Gu, Q.; Kuwajerwala, A.; Morin, S.; Jatavallabhula, K. M.; Sen, B.; Agarwal, A.; Rivera, C.; Paul, W.; Ellis, K.; Chellappa, R.; et al. 2023. Conceptgraphs: Open-vocabulary 3d scene graphs for perception and planning. *arXiv preprint arXiv:2309.16650*.

Guan, L.; Valmeekam, K.; Sreedharan, S.; and Kambhampati, S. 2023. Leveraging Pre-trained Large Language Models to Construct and Utilize World Models for Model-based Task Planning. *arXiv preprint arXiv:2305.14909*.

Hao, S.; Gu, Y.; Ma, H.; Hong, J. J.; Wang, Z.; Wang, D. Z.; and Hu, Z. 2023a. Reasoning with language model is planning with world model. *arXiv preprint arXiv:2305.14992*.

Hao, S.; Liu, T.; Wang, Z.; and Hu, Z. 2023b. ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings. *arXiv preprint arXiv:2305.11554*.

Hazra, R.; Martires, P. Z. D.; and De Raedt, L. 2023. SayCanPay: Heuristic Planning with Large Language Models using Learnable Domain Knowledge. *arXiv preprint arXiv:2308.12682*.

Howey, R.; Long, D.; and Fox, M. 2004. VAL: automatic plan validation, continuous effects and mixed initiative planning using PDDL. In *16th IEEE International Conference on Tools with Artificial Intelligence*, 294–301.

Hsieh, C.-Y.; Chen, S.-A.; Li, C.-L.; Fujii, Y.; Ratner, A.; Lee, C.-Y.; Krishna, R.; and Pfister, T. 2023. Tool documentation enables zero-shot tool-usage with large language models. *arXiv preprint arXiv:2308.00675*.

Hu, B.; Zhao, C.; Zhang, P.; Zhou, Z.; Yang, Y.; Xu, Z.; and Liu, B. 2023a. Enabling Efficient Interaction between an Algorithm Agent and an LLM: A Reinforcement Learning Approach. *arXiv preprint arXiv:2306.03604*.

Hu, H.; Lu, H.; Zhang, H.; Lam, W.; and Zhang, Y. 2023b. Chain-of-Symbol Prompting Elicits Planning in Large Langauge Models. *arXiv preprint arXiv:2305.10276*.

Hu, M.; Mu, Y.; Yu, X.; Ding, M.; Wu, S.; Shao, W.; Chen, Q.; Wang, B.; Qiao, Y.; and Luo, P. 2023c. Tree-Planner: Efficient Close-loop Task Planning with Large Language Models. *arXiv preprint arXiv:2310.08582*.

Hu, P.; Qi, J.; Li, X.; Li, H.; Wang, X.; Quan, B.; Wang, R.; and Zhou, Y. 2023d. Tree-of-mixed-thought: Combining fast and slow thinking for multi-hop visual reasoning. *arXiv preprint arXiv:2308.09658*.

Hu, Z.; and Shu, T. 2023. Language Models, Agent Models, and World Models: The LAW for Machine Reasoning and Planning. arXiv:2312.05230.

Hua, W.; Fan, L.; Li, L.; Mei, K.; Ji, J.; Ge, Y.; Hemphill, L.; and Zhang, Y. 2023. War and peace (waragent): Large language model-based multi-agent simulation of world wars. *arXiv preprint arXiv:2311.17227*.

Huang, W.; Abbeel, P.; Pathak, D.; and Mordatch, I. 2022a. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning*, 9118–9147. PMLR.

Huang, W.; Wang, C.; Zhang, R.; Li, Y.; Wu, J.; and Fei-Fei, L. 2023a. Voxposer: Composable 3d value maps for robotic manipulation with language models. *arXiv preprint arXiv:2307.05973*.

Huang, W.; Xia, F.; Shah, D.; Driess, D.; Zeng, A.; Lu, Y.; Florence, P.; Mordatch, I.; Levine, S.; Hausman, K.; et al. 2023b. Grounded decoding: Guiding text generation with grounded models for robot control. *arXiv preprint arXiv:2303.00855*.

Huang, W.; Xia, F.; Xiao, T.; Chan, H.; Liang, J.; Florence, P.; Zeng, A.; Tompson, J.; Mordatch, I.; Chebotar, Y.; et al. 2022b. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*.

Jha, S. K.; Jha, S.; Lincoln, P.; Bastian, N. D.; Velasquez, A.; Ewetz, R.; and Neema, S. 2023. Neuro Symbolic Reasoning for Planning: Counterexample Guided Inductive Synthesis using Large Language Models and Satisfiability Solving. *arXiv preprint arXiv:2309.16436*.

Joublin, F.; Ceravola, A.; Smirnov, P.; Ocker, F.; Deigmoeller, J.; Belardinelli, A.; Wang, C.; Hasler, S.; Tanneberg, D.; and Gienger, M. 2023. CoPAL: Corrective Planning of Robot Actions with Large Language Models. *arXiv preprint arXiv:2310.07263*.

Kannan, S. S.; Venkatesh, V. L.; and Min, B.-C. 2023. SMART-LLM: Smart Multi-Agent Robot Task Planning using Large Language Models. *arXiv preprint arXiv:2309.10062*.

Kant, Y.; Ramachandran, A.; Yenamandra, S.; Gilitschenski, I.; Batra, D.; Szot, A.; and Agrawal, H. 2022. Housekeep: Tidying virtual households using commonsense reasoning. In *European Conference on Computer Vision*, 355–373. Springer.

Kautz, H. A. 2022. The third AI summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine*, 43(1): 105–125.

Kelly, J.; Calderwood, A.; Wardrip-Fruin, N.; and Mateas, M. 2023. There and back again: extracting formal domains for controllable neurosymbolic story authoring. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 19, 64–74.

Kim, G.; Kim, T.; Kannan, S. S.; Venkatesh, V. L.; Kim, D.; and Min, B.-C. 2023. DynaCon: Dynamic Robot Planner with Contextual Awareness via LLMs. *arXiv preprint arXiv:2309.16031*.

Kirk, J. R.; Wray, R. E.; and Laird, J. E. 2023. Exploiting Language Models as a Source of Knowledge for Cognitive Agents. *arXiv preprint arXiv:2310.06846*.

Lenat, D.; and Marcus, G. 2023. Getting from generative ai to trustworthy ai: What llms might learn from cyc. *arXiv preprint arXiv:2308.04445*.

Li, M.; Song, F.; Yu, B.; Yu, H.; Li, Z.; Huang, F.; and Li, Y. 2023a. Api-bank: A benchmark for tool-augmented llms. *arXiv preprint arXiv:2304.08244*.

Li, Y.; Kamra, N.; Desai, R.; and Halevy, A. 2023b. Human-Centered Planning. *arXiv preprint arXiv:2311.04403*.

Lin, B. Y.; Fu, Y.; Yang, K.; Ammanabrolu, P.; Brahman, F.; Huang, S.; Bhagavatula, C.; Choi, Y.; and Ren, X. 2023a. SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks. *arXiv preprint arXiv:2305.17390*.

Lin, H.; Zala, A.; Cho, J.; and Bansal, M. 2023b. Videodirectorgpt: Consistent multi-scene video generation via llm-guided planning. *arXiv preprint arXiv:2309.15091*.

Lin, K.; Agia, C.; Migimatsu, T.; Pavone, M.; and Bohg, J. 2023c. Text2motion: From natural language instructions to feasible plans. *arXiv preprint arXiv:2303.12153*.

Liu, B.; Jiang, Y.; Zhang, X.; Liu, Q.; Zhang, S.; Biswas, J.; and Stone, P. 2023. Llm+ p: Empowering large language models with optimal planning proficiency. *arXiv preprint arXiv:2304.11477*.

Liu, Z.; Bahety, A.; and Song, S. 2023. Reflect: Summarizing robot experiences for failure explanation and correction. *arXiv preprint arXiv:2306.15724*.

Logeswaran, L.; Sohn, S.; Lyu, Y.; Liu, A. Z.; Kim, D.-K.; Shim, D.; Lee, M.; and Lee, H. 2023. Code Models are Zero-shot Precondition Reasoners. *arXiv preprint arXiv:2311.09601*.

Lu, P.; Peng, B.; Cheng, H.; Galley, M.; Chang, K.-W.; Wu, Y. N.; Zhu, S.-C.; and Gao, J. 2023a. Chameleon: Plug-and-play compositional reasoning with large language models. *arXiv preprint arXiv:2304.09842*.

Lu, Y.; Feng, W.; Zhu, W.; Xu, W.; Wang, X. E.; Eckstein, M.; and Wang, W. Y. 2022. Neuro-symbolic causal language planning with commonsense prompting. *arXiv e-prints*, arXiv–2206.

Lu, Y.; Lu, P.; Chen, Z.; Zhu, W.; Wang, X. E.; and Wang, W. Y. 2023b. Multimodal Procedural Planning via Dual Text-Image Prompting. *arXiv preprint arXiv:2305.01795*.

Luo, L.; Li, Y.-F.; Haffari, G.; and Pan, S. 2023. Reasoning on graphs: Faithful and interpretable large language model reasoning. *arXiv preprint arXiv:2310.01061*.

Mandi, Z.; Jain, S.; and Song, S. 2023. Roco: Dialectic multi-robot collaboration with large language models. *arXiv preprint arXiv:2307.04738*.

McDermott, D.; Ghallab, M.; Howe, A.; Knoblock, C.; Ram, A.; Veloso, M.; Weld, D.; and Wilkins, D. 1998. PDDL-the planning domain definition language.

Momennejad, I.; Hasanbeig, H.; Vieira, F.; Sharma, H.; Ness, R. O.; Jojic, N.; Palangi, H.; and Larson, J. 2023. Evaluating Cognitive Maps and Planning in Large Language Models with CogEval. *arXiv preprint arXiv:2309.15129*.

Naik, R.; Chandrasekaran, V.; Yuksekgonul, M.; Palangi, H.; and Nushi, B. 2023. Diversity of Thought Improves Reasoning Abilities of Large Language Models. *arXiv preprint arXiv:2310.07088*.

Nottingham, K.; Ammanabrolu, P.; Suhr, A.; Choi, Y.; Hajishirzi, H.; Singh, S.; and Fox, R. 2023. Do embodied agents dream of pixelated sheep?: Embodied decision making using language guided world modelling. *arXiv preprint arXiv:2301.12050*.

Pallagani, V.; Muppasani, B.; Murugesan, K.; Rossi, F.; Horesh, L.; Srivastava, B.; Fabiano, F.; and Loreggia, A. 2022. Plansformer: Generating symbolic plans using transformers. *arXiv preprint arXiv:2212.08681*.

Pallagani, V.; Muppasani, B.; Murugesan, K.; Rossi, F.; Srivastava, B.; Horesh, L.; Fabiano, F.; and Loreggia, A. 2023a. Understanding the Capabilities of Large Language Models for Automated Planning. *arXiv preprint arXiv:2305.16151*.

Pallagani, V.; Muppasani, B.; Srivastava, B.; Rossi, F.; Horesh, L.; Murugesan, K.; Loreggia, A.; Fabiano, F.; Joseph, R.; Kethepalli, Y.; et al. 2023b. Plansformer Tool: Demonstrating Generation of Symbolic Plans Using Transformers. In *IJCAI*, volume 2023, 7158–7162. International Joint Conferences on Artificial Intelligence.

Pan, L.; Albalak, A.; Wang, X.; and Wang, W. Y. 2023. Logic-lm: Empowering large language models with symbolic solvers for faithful logical reasoning. *arXiv preprint arXiv:2305.12295*.

Parakh, M.; Fong, A.; Simeonov, A.; Gupta, A.; Chen, T.; and Agrawal, P. 2023. Human-Assisted Continual Robot Learning with Foundation Models. *arXiv preprint arXiv:2309.14321*.

Raimondo, S.; Pal, C.; Liu, X.; Vazquez, D.; and Palacios, H. 2023. Improving Generalization in Task-oriented Dialogues with Workflows and Action Plans. *arXiv preprint arXiv:2306.01729*.

Rajvanshi, A.; Sikka, K.; Lin, X.; Lee, B.; Chiu, H.-P.; and Velasquez, A. 2023. Saynav: Grounding large language models for dynamic planning to navigation in new environments. *arXiv preprint arXiv:2309.04077*.

Raman, S. S.; Cohen, V.; Rosen, E.; Idrees, I.; Paulius, D.; and Tellex, S. 2022. Planning with large language models via corrective re-prompting. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.

Rana, K.; Haviland, J.; Garg, S.; Abou-Chakra, J.; Reid, I.; and Suenderhauf, N. 2023. Sayplan: Grounding large language models using 3d scene graphs for scalable task planning. *arXiv preprint arXiv:2307.06135*.

Ren, A. Z.; Dixit, A.; Bodrova, A.; Singh, S.; Tu, S.; Brown, N.; Xu, P.; Takayama, L.; Xia, F.; Varley, J.; et al. 2023. Robots that ask for help: Uncertainty alignment for large language model planners. *arXiv preprint arXiv:2307.01928*.

Rossetti, N.; Tummolo, M.; Gerevini, A.; Putelli, L.; Serina, I.; Chiari, M.; and Olivato, M. 2024. Learning General Policies for Planning through GPT Models. In *34th International Conference on Automated Planning and Scheduling*.

Ruan, J.; Chen, Y.; Zhang, B.; Xu, Z.; Bao, T.; Du, G.; Shi, S.; Mao, H.; Zeng, X.; and Zhao, R. 2023. Tptu: Task planning and tool usage of large language model-based ai agents. *arXiv preprint arXiv:2308.03427*.

Russell, S.; and Norvig, P. 2003. *Artificial Intelligence, A Modern Approach. Second Edition*.

Sakib, M. S.; and Sun, Y. 2023. From Cooking Recipes to Robot Task Trees–Improving Planning Correctness and Task Efficiency by Leveraging LLMs with a Knowledge Network. *arXiv preprint arXiv:2309.09181*.

Sarkisyan, C.; Korchemnyi, A.; Kovalev, A. K.; and Panov, A. I. 2023. Evaluation of Pretrained Large Language Models in Embodied Planning Tasks. In *International Conference on Artificial General Intelligence*, 222–232. Springer.

Sermanet, P.; Ding, T.; Zhao, J.; Xia, F.; Dwibedi, D.; Gopalakrishnan, K.; Chan, C.; Dulac-Arnold, G.; Maddineni, S.; Joshi, N. J.; et al. 2023. RoboVQA: Multimodal Long-Horizon Reasoning for Robotics. *arXiv preprint arXiv:2311.00899*.

Shah, D.; Equi, M.; Osinski, B.; Xia, F.; Ichter, B.; and Levine, S. 2023. Navigation with large language models: Semantic guesswork as a heuristic for planning. *arXiv preprint arXiv:2310.10103*.

Shen, Y.; Song, K.; Tan, X.; Li, D.; Lu, W.; and Zhuang, Y. 2023. Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*.

Shirai, K.; Beltran-Hernandez, C. C.; Hamaya, M.; Hashimoto, A.; Tanaka, S.; Kawaharazuka, K.; Tanaka, K.; Ushiku, Y.; and Mori, S. 2023. Vision-Language Interpreter for Robot Task Planning. *arXiv preprint arXiv:2311.00967*.

Silver, T.; Dan, S.; Srinivas, K.; Tenenbaum, J. B.; Kaelbling, L. P.; and Katz, M. 2023. Generalized Planning in PDDL Domains with Pretrained Large Language Models. *arXiv preprint arXiv:2305.11014*.

Silver, T.; Hariprasad, V.; Shuttleworth, R. S.; Kumar, N.; Lozano-Pérez, T.; and Kaelbling, L. P. 2022. PDDL planning with pretrained large language models. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.

Singh, I.; Blukis, V.; Mousavian, A.; Goyal, A.; Xu, D.; Tremblay, J.; Fox, D.; Thomason, J.; and Garg, A. 2023. ProgPrompt: program generation for situated robot task planning using large language models. *Autonomous Robots*, 1–14.

Song, C. H.; Wu, J.; Washington, C.; Sadler, B. M.; Chao, W.-L.; and Su, Y. 2023. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2998–3009.

Sumers, T.; Yao, S.; Narasimhan, K.; and Griffiths, T. L. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*.

Sun, H.; Zhuang, Y.; Kong, L.; Dai, B.; and Zhang, C. 2023. AdaPlanner: Adaptive Planning from Feedback with Language Models. *arXiv preprint arXiv:2305.16653*.

Tang, X.; Zheng, Z.; Li, J.; Meng, F.; Zhu, S.-C.; Liang, Y.; and Zhang, M. 2023. Large Language Models are In-Context Semantic Reasoners rather than Symbolic Reasoners. *arXiv preprint arXiv:2305.14825*.

Valmeekam, K.; Marquez, M.; and Kambhampati, S. 2023. Can Large Language Models Really Improve by Self-critiquing Their Own Plans? *arXiv preprint arXiv:2310.08118*.

Valmeekam, K.; Marquez, M.; Olmo, A.; Sreedharan, S.; and Kambhampati, S. 2023a. PlanBench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Valmeekam, K.; Olmo, A.; Sreedharan, S.; and Kambhampati, S. 2022. Large Language Models Still Can't Plan (A Benchmark for LLMs on Planning and Reasoning about Change). *arXiv preprint arXiv:2206.10498*.

Valmeekam, K.; Sreedharan, S.; Marquez, M.; Olmo, A.; and Kambhampati, S. 2023b. On the planning abilities of large language models (a critical investigation with a proposed benchmark). *arXiv preprint arXiv:2302.06706*.

Wang, J.; Tong, J.; Tan, K.; Vorobeychik, Y.; and Kantaros, Y. 2023a. Conformal Temporal Logic Planning using Large Language Models: Knowing When to Do What and When to Ask for Help. *arXiv preprint arXiv:2309.10092*.

Wang, L.; Xu, W.; Lan, Y.; Hu, Z.; Lan, Y.; Lee, R. K.-W.; and Lim, E.-P. 2023b. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models. *arXiv preprint arXiv:2305.04091*.

Wang, X.; Caccia, L.; Ostapenko, O.; Yuan, X.; and Sordoni, A. 2023c. Guiding language model reasoning with planning tokens. *arXiv preprint arXiv:2310.05707*.

Wang, Z.; Cai, S.; Liu, A.; Ma, X.; and Liang, Y. 2023d. Describe, explain, plan and select: Interactive planning with large language models enables open-world multi-task agents. *arXiv preprint arXiv:2302.01560*.

Webb, T.; Mondal, S. S.; Wang, C.; Krabach, B.; and Momennejad, I. 2023. A Prefrontal Cortex-inspired Architecture for Planning in Large Language Models. *arXiv preprint arXiv:2310.00194*.

Wei, L.; He, Z.; Zhao, H.; and Yao, Q. 2023. Unleashing the Power of Graph Learning through LLM-based Autonomous Agents. *arXiv preprint arXiv:2309.04565*.

Wong, L.; Grand, G.; Lew, A. K.; Goodman, N. D.; Mansinghka, V. K.; Andreas, J.; and Tenenbaum, J. B. 2023. From Word Models to World Models: Translating from Natural Language to the Probabilistic Language of Thought. *arXiv preprint arXiv:2306.12672*.

Wu, Y.; Min, S. Y.; Bisk, Y.; Salakhutdinov, R.; Azaria, A.; Li, Y.; Mitchell, T.; and Prabhumoye, S. 2023a. Plan, Eliminate, and Track–Language Models are Good Teachers for Embodied Agents. *arXiv preprint arXiv:2305.02412*.

Wu, Z.; Ai, B.; and Hsu, D. 2023. Integrating Common Sense and Planning with Large Language Models for Room Tidying. In *RSS 2023 Workshop on Learning for Task and Motion Planning*.

Wu, Z.; Wang, Z.; Xu, X.; Lu, J.; and Yan, H. 2023b. Embodied task planning with large language models. *arXiv preprint arXiv:2307.01848*.

Xie, Y.; Yu, C.; Zhu, T.; Bai, J.; Gong, Z.; and Soh, H. 2023. Translating natural language to planning goals with large-language models. *arXiv preprint arXiv:2302.05128*.

Xu, B.; Liu, X.; Shen, H.; Han, Z.; Li, Y.; Yue, M.; Peng, Z.; Liu, Y.; Yao, Z.; and Xu, D. 2023a. Gentopia: A collaborative platform for tool-augmented llms. *arXiv preprint arXiv:2308.04030*.

Xu, L. 1995. Case based reasoning. *IEEE Potentials*, 13(5): 10–13.

Xu, M.; Huang, P.; Yu, W.; Liu, S.; Zhang, X.; Niu, Y.; Zhang, T.; Xia, F.; Tan, J.; and Zhao, D. 2023b. Creative Robot Tool Use with Large Language Models. *arXiv preprint arXiv:2310.13065*.

Yang, J.; Chen, X.; Qian, S.; Madaan, N.; Iyengar, M.; Fouhey, D. F.; and Chai, J. 2023a. LLM-Grounder: Open-Vocabulary 3D Visual Grounding with Large Language Model as an Agent. *arXiv preprint arXiv:2309.12311*.

Yang, R.; Hou, M.; Wang, J.; and Zhang, F. 2023b. OceanChat: Piloting Autonomous Underwater Vehicles in Natural Language. *arXiv preprint arXiv:2309.16052*.

Yang, Y.; Gaglione, J.-R.; and Topcu, U. 2022. Learning Automata-Based Task Knowledge Representation from Large-Scale Generative Language Models. *arXiv preprint arXiv:2212.01944*.

Yang, Y.; and Tomar, A. 2023. On the Planning, Search, and Memorization Capabilities of Large Language Models. *arXiv preprint arXiv:2309.01868*.

Yang, Z. 2023. *Neuro-Symbolic AI Approaches to Enhance Deep Neural Networks with Logical Reasoning and Knowledge Integration*. Ph.D. thesis, Arizona State University.

Yang, Z.; Ishay, A.; and Lee, J. 2023. Coupling Large Language Models with Logic Programming for Robust and General Reasoning from Text. *arXiv preprint arXiv:2307.07696*.

Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T. L.; Cao, Y.; and Narasimhan, K. 2023. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*.

Yoneda, T.; Fang, J.; Li, P.; Zhang, H.; Jiang, T.; Lin, S.; Picker, B.; Yunis, D.; Mei, H.; and Walter, M. R. 2023. Statler: State-maintaining language models for embodied reasoning. *arXiv preprint arXiv:2306.17840*.

You, W.; Wu, W.; Liang, Y.; Mao, S.; Wu, C.; Cao, M.; Cai, Y.; Guo, Y.; Xia, Y.; Wei, F.; et al. 2023. EIPE-text: Evaluation-Guided Iterative Plan Extraction for Long-Form Narrative Text Generation. *arXiv preprint arXiv:2310.08185*.

Yuan, H.; Zhang, C.; Wang, H.; Xie, F.; Cai, P.; Dong, H.; and Lu, Z. 2023a. Plan4mc: Skill reinforcement learning and planning for open-world minecraft tasks. *arXiv preprint arXiv:2303.16563*.

Yuan, S.; Chen, J.; Fu, Z.; Ge, X.; Shah, S.; Jankowski, C. R.; Yang, D.; and Xiao, Y. 2023b. Distilling Script Knowledge from Large Language Models for Constrained Language Planning. *arXiv preprint arXiv:2305.05252*.

Zelikman, E.; Huang, Q.; Poesia, G.; Goodman, N.; and Haber, N. 2023. Parsel: Algorithmic Reasoning with Language Models by Composing Decompositions. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Zhang, B.; and Soh, H. 2023. Large language models as zero-shot human models for human-robot interaction. *arXiv preprint arXiv:2303.03548*.

Zhang, C.; Liu, L.; Wang, J.; Wang, C.; Sun, X.; Wang, H.; and Cai, M. 2023a. Prefer: Prompt ensemble learning via feedback-reflect-refine. *arXiv preprint arXiv:2308.12033*.

Zhang, F.; Jin, K.; and Zhuo, H. H. 2023. Planning with Logical Graph-based Language Model for Instruction Generation. arXiv:2308.13782.

Zhang, H.; Du, W.; Shan, J.; Zhou, Q.; Du, Y.; Tenenbaum, J. B.; Shu, T.; and Gan, C. 2023b. Building cooperative embodied agents modularly with large language models. *arXiv preprint arXiv:2307.02485*.

Zhang, J.; Zhang, J.; Pertsch, K.; Liu, Z.; Ren, X.; Chang, M.; Sun, S.-H.; and Lim, J. J. 2023c. Bootstrap your own skills: Learning to solve new tasks with large language model guidance. *arXiv preprint arXiv:2310.10021*.

Zhao, Z.; Lee, W. S.; and Hsu, D. 2023. Large Language Models as Commonsense Knowledge for Large-Scale Task Planning. *arXiv preprint arXiv:2305.14078*.

Zheng, L.; Chiang, W.-L.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Li, D.; Xing, E. P.; Zhang, H.; Gonzalez, J. E.; and Stoica, I. 2023a. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. arXiv:2306.05685.

Zheng, S.; Liu, J.; Feng, Y.; and Lu, Z. 2023b. Steve-Eye: Equipping LLM-based Embodied Agents with Visual Perception in Open Worlds. *arXiv preprint arXiv:2310.13255*.

Zhou, Z.; Song, J.; Yao, K.; Shu, Z.; and Ma, L. 2023. ISR-LLM: Iterative Self-Refined Large Language Model for Long-Horizon Sequential Task Planning. *arXiv preprint arXiv:2308.13724*.