From Language to Action: Employing Foundation Models in Autonomous Robots

Anonymous ACL submission

Abstract

Foundation models have demonstrated remarkable capabilities in natural language processing tasks, generating interest in their potential for robotic applications. However, the existing literature lacks a transparent and comprehensive synthesis of these advancements. This paper utilizes the PRISMA framework to systematically review and explore the integration of foundation models in robotic applications. Through an in-depth analysis of 76 studies, we investigate current trends in models, modalities, and experimental methods. Additionally, this study maps the state-of-the-art applications of foundation models in robotics tasks, and illustrate how these tasks are interconnected. Synthesizing these findings, we identified key challenges and future direction. This study establishes a benchmark and offers insights into future research directions for developing safe and autonomous embodied foundation models. All data, and findings are available on the project repository [1].

1 Introduction

Foundation models are defined as large-scale Artificial Intelligence (AI) models trained on an extensive and internet-scale dataset, capable of generalizing knowledge across a wide range of tasks. These models utilize massive datasets in a self-supervised manner to learn from unannoted data, allowing them to be adapted to various downstream tasks (Bommasani et al., 2021). Generalizing across diverse tasks without tasks-specific fine-tunning in models, such as GPT-4 (Achiam et al., 2023) Llama-2 (Touvron et al., 2023) Gemini (Anil et al., 2023) Claude (Anthropic, 2023), have significantly advanced the natural language processing (NLP) field. Such strengths along with their adaptability and ability to process multi-modal data (text, image, sound) have drawn the attention of researchers in various domains, ranging from the medical field (Cho et al., 2023) to robotics (Xiao et al., 2023) to bring cognitive capabilities of these models to physical world applications.

To achieve a degree of autonomy in physical world, embodied agents or robots have been utilized from many years ago (Smithers, 1997). There are generally two broad solution categories for automating these embodied agents: (1) preprogramming robots for specific scenarios; (2) teleoperating robots to leverage human cognitive abilities (Saidi et al., 2016). The first category already employed AI paradigms, such as reinforcement learning (Delgado and Oyedele, 2022) and deep learning (Karoly et al., 2021), to automate specific labor-intensive and repetitive tasks (Bruun et al., 2022; Yu et al., 2009). While these robots can deliver satisfactory precision in designated tasks, their adaptability and generalizability are often limited due to training on narrowly focused datasets designed for specific tasks. Consequently, manual adjustments may be necessary to accommodate even minor task variations in physical world applications (Cully et al., 2015). In contrast, the second category involves tele-operated robots, which can be remotely operated by experts, allowing them to adapt to various tasks without the need for manual reprogramming. However, their dependency on human operators has limited their performance and productivity. For example, even slight connection delays can significantly impede robot performance in extraterrestrial physical worlds (Seo et al., 2024).

On the other hand, foundation models are trained on vast amounts of data to exhibit adaptability, generalizability, and overall performance across a variety of domains (Chang et al., 2023). This intrinsic feature can be seen as a solution to move embodied agents and robots to a higher level of autonomy for physical world applications. Consequently, this study aims to: (1) systematically explore the current state of the art of tools, methods,

[1] will be available in the final version

and applications of foundation models in robotic applications; (2) investigate how foundation models have impacted the cooperation of cognitive and acting tasks in physical environments; (3) identify current challenges and provide future directions for future embodied foundation models. Therefore, this study can serve as a benchmark for other researchers to track progress toward future safe and fully autonomous embodied agents.

2 Autonomous Robot Components

Figure 1 illustrates essential components of a robot operating within a physical world. Autonomous robots are comprised of two main platforms: (1) the deliberation platform; and (2) the execution platform. The execution platform, which is influenced by the robot's morphology, includes various actuators, motors, sensors, end effectors, and manipulators. Developments in this platform are beyond the scope of this study, as our primary focus is on the deliberation platform. This platform is responsible for receiving objectives and percepts (mostly from various sensors), processing them, and generating actionable commands or communication signals.

The deliberation platform employs two main modules: (1) the cognitive module, which is responsible for all cognitive processes in robots; and (2) the acting module, which translates cognitive outputs into fine-grained actionable commands. Reasoning is the highest-level cognitive process, inferring new information from existing signals. Mid-level processes include planning, which involves decision sequences to achieve goals, and decision-making, which selects actions based on percepts and predefined criteria. Human-robot interaction enables communication through speech recognition, natural language processing, and understanding gestures or facial expressions. Perception involves processing environmental information, including object recognition, scene understanding, SLAM, and gesture recognition. The acting module controls actuators for executing actions, navigating through environments with path planning and obstacle avoidance, and manipulating objects.

2.1 Related Studies

To date of drafting this manuscript, three studies have surveyed the application of foundation models in robotics. The first review paper by Firoozi et al. 2023 surveyed the application of foundation models in robotics with an emphasis on future challenges and opportunities. The second review paper by Xiao et al. 2023 explored existing studies focused on robot learning using foundation models to identify potential future areas. In the third review, Hu et al. 2023 examined different studies relevant to foundation models and investigated how their application could be adapted to the robotics field.

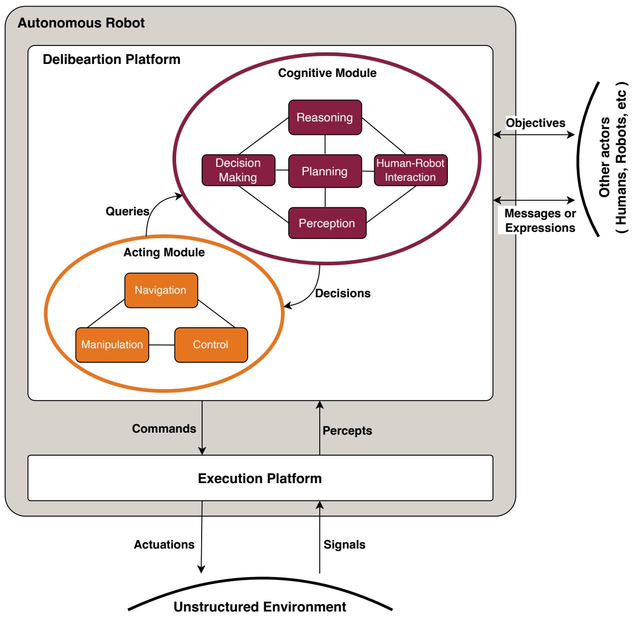

Figure 1: Conceptual view of robot in physical world applications

These studies lack a transparent and reproducible approach for categorizing their findings and providing insights for future research. While they do categorize studies, they fail to accurately highlight the significance of each field, making it difficult to compare and analyze which applications need more attention from researchers. This method also falls short in identifying subtle research gaps that are not apparent through narrative categorization. Consequently, there is a lack of an objective benchmark in the field to track progress and ensure that studies are advancing safely and aligning with our goals. This study is distinguished from previous ones for the following reasons: (1) Our study builds an objective picture of the current state-of-the-art in employing foundation models for robotic applications. We map the impact of these models across different cognitive and acting tasks and explore the correlations between them. (2) The provided current state-of-the-art are synthesized to identify new challenges and potential future research directions, paving the way for a safe and autonomous future in the field. (3) Our study employs a transparent and reproducible methodology, aiming to establish a clear and objective benchmark for future research.

3 Methodology

A systematic review approach has been selected for this study based on PRISMA framework (Page et al., 2021) to explore the embodiment of foundation models in physical worlds through robots.

3.1 Databases

Web of Science (WoS) and Scopus are two comprehensive databases serving as major tools for systematic review in the field of science, technology, engineering and mathematics (STEM) (Kandall, 2017; Visser et al., 2020). In addition to these, ArXiv is selected as the main source for preprint studies within the scope of our study because: (1) it serves as one of the main sources of studies related to foundation models from 2018 up to now (Gusenbauer and Haddaway, 2020); (2) it helps us to cover emerging ideas that are not yet published in journals due to the long process of publishing (Movva et al., 2023).

3.2 Search query

Query-based search is one of the most fundamental methods for identifying relevant studies in a field of research (Chen and Song, 2019). To maximize the potential of identifying relevant studies within our scope, we constructed two word-family blocks, containing keywords relevant to our targeted studies (see Figure 2). Within these blocks, keywords are connected with "OR" command to maximize the likelihood of retrieving relevant studies. Among these blocks, the word-family block for foundation models (left block) is connected with "AND" command to the word-family block for robotics (right block). Linking the left block with the right block generates a search query suitable for exploring the application of foundation models in robotics for physical world applications.

3.3 Screening

Figure 3 illustrates the process of identifying relevant studies for this survey. It should be noted that the number of studies at each step is dependent on the date of drafting this manuscript was drafted (March 2024). Initially, the search query was applied to identified databases, followed by the exclusion of duplicate studies. Subsequently, several eligibility refinements, such as language, date, and study types were made to the search outputs to align them more closely with the study's scope. Noting that the first versions of foundation models emerged in 2018, we restricted the identified studies to the time frame of 2018 to 2024. In the next step, we established two set of screening criteria to ensure that the identified studies are relevant to our scope.

4 Results

This section aims to provide an objective picture of the current state-of-the-art in the applicability of foundation models for automating tasks in physical world using robots. To achieve this goal, all identified studies were subjected to a comprehensive whole-text content analysis. We extract a set of 20 features to have a detailed and complete overview of the recent trend.

These features can be categorized to two 10 feature groups: (1) general features: authors, title, published year, source title, DOI, link to paper, author affiliations, abstract, author and index keywords; (2) specific features: applications, foundation model use, applied tasks, domain, study objective, robot morphology, evaluation method, modalities, transformer architecture, and open source status. Due to the limited space, we present a subset of the features in the main paper while description and details of all other features are available under the open source licence [2].

4.1 Foundation model usage trends

This section investigates the frequency of utilizing foundation models for robotic and physical world applications. As seen in Figure 4, the integration of foundation models into robotics is dominated by GPT-Based models, which account for over 44% of foundation model usage. GPT-3.5 is the most frequently used LLM, highlighting its applicability and ease of use. Although GPT-4 is located in third usage place, it should be noted that the usage of GPT-4 is rapidly growing.

Another interesting finding is that CLIP model is the most frequent used models among Visual Language Models (VLMs) and second place among all foundation model usages in robotic applications. CLIP is primarily utilized for bridging similarities between text, as the first source of receiving language instructions, and images, as the primary means of understanding environments. It has been

[2] will be available in the final version

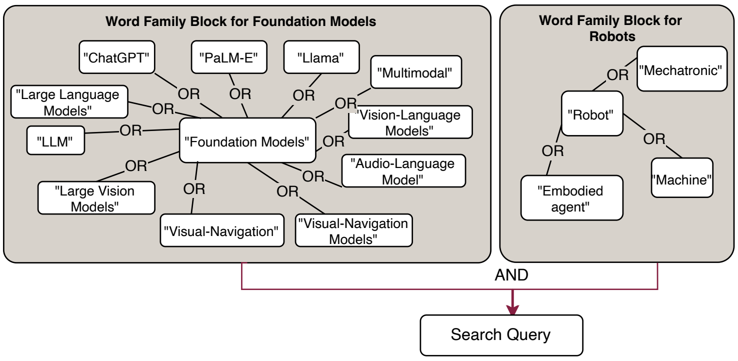

Figure 2: Process of building search queries

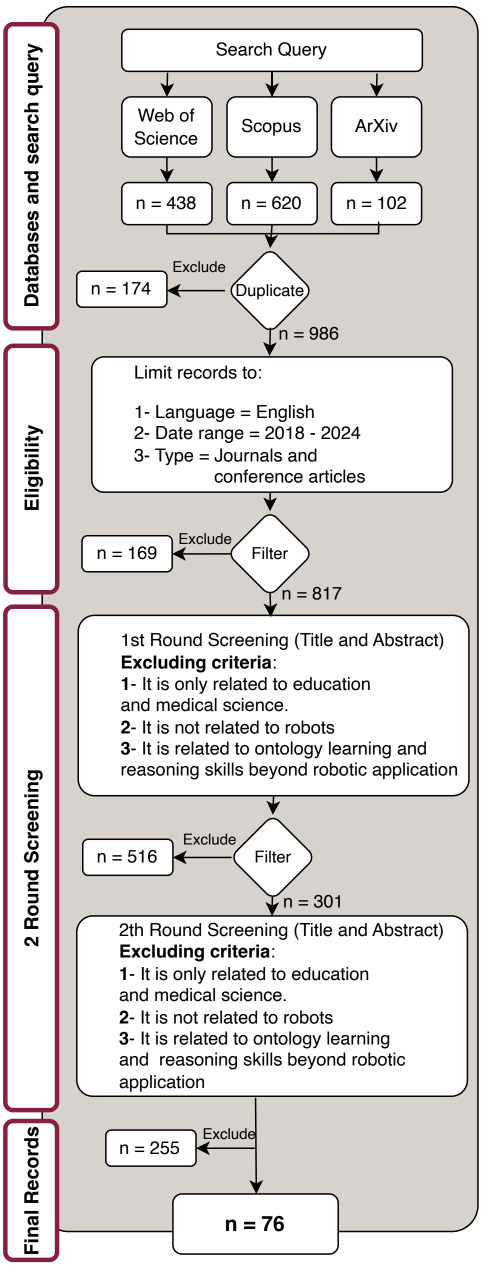

Figure 3: Process of identifying relevant records (PRISMA)

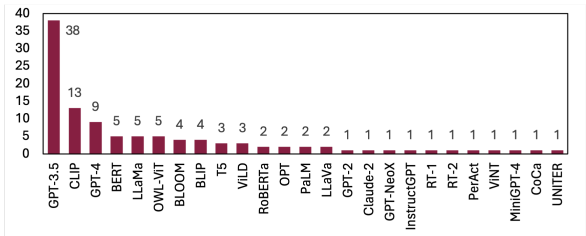

Figure 4: Frequency of using different foundation models in robotic field

used in various studies to perform Vision-Language Navigation (VLN) related tasks (Lan et al., 2023; Lin et al., 2022), as well as other manipulation tasks (Cui et al., 2022; Liao et al., 2023; Shridhar et al., 2021), and even high-level recognition tasks, such as reasoning (Kamath et al., 2023). As seen in Figure 4, the frequency of remaining models is five or fewer,

4.2 Modalities

As mentioned in Table 1, more than half of the identified studies utilize only text for developing their use cases. Although it indicates the early stages of studies in this field, 31% multimodal text and image models indicate the move toward more multimodal models. However, the number of studies using other modalities such as 3d and audio data is very limited. This lack of diversity in modalities may hinder the development of more comprehensive and robust robotic systems capable of perceiving and interacting with the real world, which is inherently multimodal.

4.3 Experiment method

Within the final records, %42 studies tested their findings through implementation in real-world experiments (see Table 2. However, this amount for

Status	Number	Percentage
Text	39	%51
Image	5	%6
Image and text	24	%31
3D data	1	%1
Audio	1	%1
Not available	1	%10

Table 1: Modalities of foundation models utilized in identified studies.

Experiment method	Number	Percentage
Real-world	32	%42
Simulation	10	%13
Dataset	7	%9
Not available	27	%36

Table 2: Experiment methods of foundation models utilized in identified studies.

simulation and dataset experiments are respectively %13 and %9. This indicates a gap that there is still a need for more comprehensive and diverse evaluation methods. Moreover, a considerable amount of studies (%36) are conceptual and doesn't validate their findings through experiments. Therefore, more studies are needed in this field to bridge the gap between theory and practice, and to thoroughly evaluate the performance and limitations of foundation models in realistic robotic applications.

4.4 Current State of the Art: Application of Foundation Models in Robotics

This section aims to provide a map of the current state of the use of different foundation models in robotic tasks. To achieve this, the identified records were labeled based on the foundation models used and the specific tasks to which these models were applied (foundation model use and applied tasks features). Figure 5 illustrates the flow of applying different foundation models for robotic tasks. This figure is organized across four analytical layers: foundation models, their categories, and categories of robotic tasks, and the specific robotic tasks.

Foundation model categories: Within the foundation model categories, Large Language Models (LLMs) contributed to 69% of foundation models utilized for robotic applications, indicating that most studies are exploring the text modalities and capabilities of this category for addressing classic challenges in robotic domains. For example, some studies utilize the capabilities of these foundation models in understanding language and coding to generate robotics execution codes in industries (Fan et al., 2024; Yoshikawa et al., 2023). VLMs also contributed another 20% of foundation model applications in robotic tasks, helping to bridge language instructions with vision perception in various studies (Kawaharazuka et al., 2023). However, less attention has been given to the application of LVMs (%7) in the robotic domain, where further studies are needed. Moreover, a few studies have gone beyond text or image-based foundation models by creating robot transformers (Brohan et al., 2023; Stone et al., 2023), yet more studies, such as MiniGPT-3D (Tang et al., 2024), are felt necessary to build 3D foundation models as they can contribute more significantly to robot-specific tasks that require direct interaction with the 3D world.

Planning and perception tasks: When it comes to robotic tasks, foundation models are primarily (72%) utilized for cognitive tasks rather than acting tasks (28%). Within the cognitive domain, perception and planning are most common goal of using foundation models in many identified records. For example, studies utilized capabilities of ChatGPT in understanding text to change the traditional method of robot planning, by generating behavior-tree (Cao and Lee, 2023) or considering the current state of robots in plan generation (Xie et al., 2023). Furthermore, some studies focused on providing robots with better perception by utilizing foundation models in complex robotic tasks, such as scene anomaly detection (Obinata et al., 2023).

Human-Robot Interaction (HRI) is another important use case of foundation models in the robotics field. The use of foundation models in HRI can be categorized into three main streams. First, some studies utilize LLMs to improve HRI through better extraction of machine-understandable information from human instructions (Bimbatti et al., 2023; Tabone and Winter, 2023). Another group of studies uses foundation models to understand public perceptions toward robots (Brandtzaeg et al., 2023; Jangjarat et al., 2023; **?**). The last group applies the capabilities of LLMs to generate human-like text to respond to humans and improve trust between humans and robots (Mishra et al., 2023; Ye et al., 2023; Sevilla-Salcedo et al., 2023).

Reasoning and decision-making tasks: In terms of reasoning, one mainstream application is the use of foundation models for providing commonsense knowledge to robots (Jain et al., 2023; Zhou et al., 2023b). Commonsense reasoning is

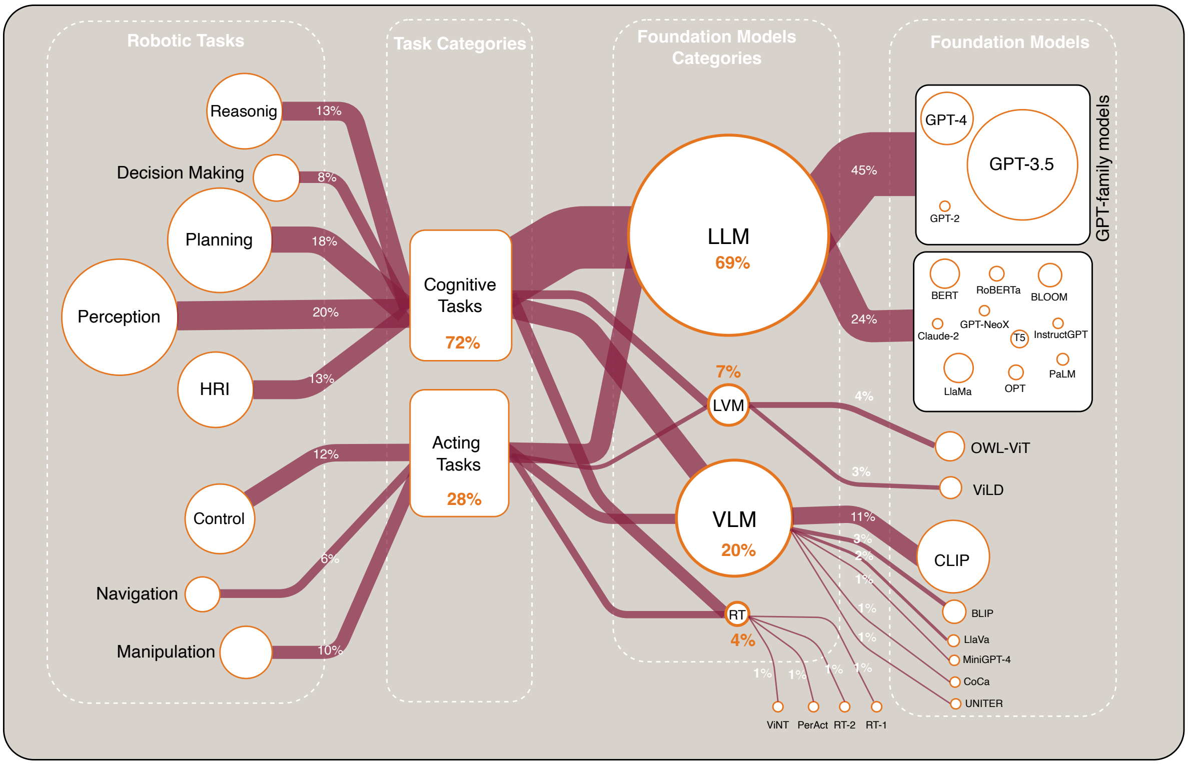

Figure 5: Flow diagram of foundation model applications in robotic tasks

a hard task for machines but it is crucial in many tasks. For example, Krause and Stolzenburg 2024 utilized LLM commonsense reasoning capabilities in the field of question answering (QA), which is one of the most important tasks of NLP. Ocker et al. 2023 found that LLMs are not sufficient enough on their own to provide commonsense reasoning but they are effective in synergy with formal knowledge representations. On the other hand, few studies investigate the decision-making abilities of LLMs in connection with different robotic tasks, such as planning (Ouyang and Li) and manipulation (Lew et al., 2023).

Control, manipulation, and navigation: Beyond cognitive tasks, the capabilities of foundation models in acting tasks are less explored. For example, some studies use language understanding of LLMs as a translation module between human and robot for controlling the simple motion of robots (Tanaka and Katsura, 2023; Kawaharazuka et al., 2023). Some other studies are providing innovative frameworks for improving spatial reasoning required in LLMs for robotic manipulation tasks (Shridhar et al., 2021; Jin et al., 2023). Navigation is another challenging tasks that recent foundation models are used to allow researchers to have semantic reasoning and go beyond conventional map-based systems (Gadre et al., 2022; Yu et al., 2023). Despite these examples, acting tasks are usually come with other cognitive tasks such as planning, and perception. As a result, a network of connection between these tasks help us to achieve better interpretation of foundation model capabilities.

4.5 Robotic Task Integration

Robotic cognitive and acting tasks are utilized in studies in an interconnected manner to automate specific tasks. Accordingly, most identified records employ foundation models across a variety of cognitive and acting tasks in a interconnected manner to evaluate and validate their research. Consequently, there is a need for a network diagram that shows how foundation models are used to interconnect different robotic tasks. Figure 6 illustrates the co-occurrence network of robotic tasks, where cognitive and acting tasks are represented as nodes. The edges between nodes represent the co-occurrence of two tasks within a single study. The size of each node is proportionate to the number of its connections, indicating that larger nodes are more frequently utilized in conjunction with other tasks in studies. The thickness of the edges indicates the frequency of concurrent task usage in the studies.

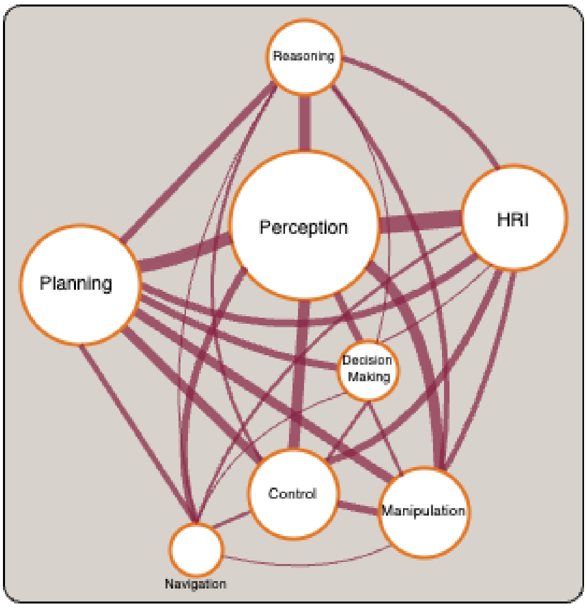

Figure 6: Co-occurrence network of robotic tasks using foundation models

As illustrated in Figure 6, the most significant connection is the use of foundation models for HRI and Perception. This finding, coupled with the dominance of LLMs in foundation models, indicates that most studies leverage the text analytical capabilities of foundation models to extract both defined and undefined information for other significant robotics tasks, including planning, control, and manipulation.

The navigation node is smaller than other nodes, indicating that navigation tasks less frequently co-occur with other robotic tasks. This suggests that the majority of the field is interested in validating the capabilities of foundation models in cognitive tasks, and some acting tasks such as control and manipulation, rather than incorporating the complexity of moving in a 3D environment into their studies. Another interesting finding is that all edges leading to the decision-making node are thin, which indicates that this task is also overlooked in many studies. Despite the small size of the reasoning task node, there is a considerable connection between this node and the perception node. This represents a major category within this field, which involves utilizing reasoning capabilities to perceive situations where only a small amount of information is available, such as unseen scenes and undefined events (Ocker et al., 2023; Ren et al., 2023; Zhang et al., 2024)

5 Discussion: Challenges and future prospects

5.1 Situated Reasoning

Currently, more studies are focused on robotic cognitive tasks (see Figure 5), which can be attributed to the fact that the current architecture of most foundation models is designed for sequential tokens, making them better suited for cognitive tasks rather than acting tasks that require extensive situated reasoning and direct interaction with 3D data. One of the main challenges in utilizing foundation models for acting tasks is the scarcity of 3D data compared to text and image data. A potential solution to this challenge can be the use of Digital Twins as a source for training foundation models on 3D data.

5.2 Physical Laws

Furthermore, the generative models, such as Sora (Liu et al., 2024), can be leveraged to create simulations of real-world environments, providing a rich source of data for training and testing. However, a significant obstacle in expanding generative models for creating simulations is their current limitation in accurately modeling physical laws, such as gravity, collisions, and other laws that are crucial for realistic simulations and interactions with the physical world. Addressing this challenge is pivotal for enabling foundation models to reason effectively about the complex dynamics and constraints of the physical world.

5.3 Hallucination

A primary issue toward effective integration of robots and foundation models is the tendency of these models to "hallucinate," meaning they sometime generate outputs that are factually incorrect, logically inconsistent, or physically infeasible. This uncertainty becomes particularly critical when robots are expected to perform a broader range of general tasks in 3D environments, especially those rarely encountered in their mostly textual and image data. Despite retrieval-based and other related path toward addressing this issue, some studies seek methods that enable LLMs to ask for help in uncertain situations (Ren et al., 2023).

5.4 Error Handling

Furthermore, this uncertainty challenge can lead to error handling due to various potential to robot action failures. These failures include: (1) execution failures, where the model understands the task and environment correctly but fails to achieve the expected outcome; (2) planning failures, where the model generates an incorrect or infeasible sequence of actions despite comprehending the task and environment; and (3) comprehension failures, where the model misinterprets the context of the environment or task. To address these issues, several approaches have been proposed. Prompt engineering methods allows the model to prompt itself with the output plan and the latest environment observations for potential corrections. Additionally, incorporating models with enhanced situated reasoning can provide more accurate predictions of robot capabilities in complex environments. Another effective strategy is leveraging human feedback, which can resolve various types of errors.

5.5 Model Biases

Recent studies highlight different biases in GPT-family models (Rutinowski et al., 2023; Sinha, 2023; Toro, 2023). Considering that currently most studies use GPT-based models (refer to Section 4.1), further research is needed to investigate the biases of using single models or a combination of different agents in more sensitive tasks, such as human-robot interaction or decision-making. Identifying these biases is a critical challenge that is crucial to tackle to ensure the safe and reliable integration of foundation models into robotic systems, especially in applications involving direct interaction with humans or decision-making processes.

5.6 Ethical considerations

While a comprehensive discussion of ethics goes beyond the scope of this study and requires extensive exploration of various ethical frameworks, it is essential to encourage more researchers to engage with this sensitive area. Key ethical considerations include privacy, safety, responsibility, and the moral behavior of robots, each of which warrants thorough examination. As a potential approach to addressing these issues, (Zhou et al., 2023a) have proposed a framework that equips foundation models with the capability for moral reasoning, drawing on diverse ethical theories. Such research is appropriate first-step as it advances the preparation of robots for deeper integration into human-centric environments, ensuring their actions are guided by sound ethical principles.

5.7 Toward unstructured environment

Most studies have tested the integration of foundation models in organized and structured environments, such as housing settings. However, unstructured environments are in greater need of foundation model capabilities due to the limitations of traditional hard-coded approaches that are unsuitable for these settings. The flexibility and generalizability inherent in foundation models can significantly enhance performance and adoption in such complex environments. Nonetheless, there are challenges in this endeavor. Unstructured environments are difficult for real-world testing applications, and we currently lack a simulation solution that accurately represents the dynamic events and unpredictability of these settings. A crucial first step is to systematically identify inherent features in unstructured environment tasks that hinder robotic adoption. For instance, future studies can explore how commonsense reasoning in foundation models can aid robot decision-making in situations where information is highly dynamic or scarce.

6 Conclusion

The integration of foundation models into robotics is an emerging field with significant potential for enabling advanced cognitive and acting capabilities in physical world applications. While the current research landscape is dominated by leveraging the language understanding abilities of LLMs, there is a growing interest in exploring multimodal and 3D foundation models for more comprehensive scene understanding and situated reasoning. However, several key challenges need to be addressed, including scarcity of 3D data, improving the modeling of physical laws in simulations, mitigating hallucinations, developing robust error-handling strategies, and addressing ethical concerns surrounding the deployment of embodied agents. Overcoming these hurdles will be crucial for realizing the vision of safe and fully autonomous embodied foundation models capable of generalizing across a wide range of unstructured environments and tasks.

7 Limitation

Despite the contributions of this study as discussed before; all research studies have limitations, and the present attempt is no exception to this rule. The survey process only considered studies in English, and used a particular set of keywords for searching. Besides, the screening process of core studies can be considered subjective in nature, although the process was performed three separate times to minimize the error. In addition, all analyses are based on the data retrieved from WoS, Scopus, and Arxiv databases. Therefore, the findings may not fully reflect the entire available efforts and studies in the field.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mo Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2023. Gpt-4 technical report. *ArXiv*.

Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Slav Petrov, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, Paul R. Barham, Tom Hennigan, Benjamin Lee, Fabio Viola, Malcolm Reynolds, Yuanzhong Xu, Ryan Doherty, Eli Collins, Clemens Meyer, Eliza Rutherford, Erica Moreira, Kareem Ayoub, Megha Goel, George Tucker, Enrique Piqueras, Maxim Krikun, Iain Barr, Nikolay Savinov, Ivo Danihelka, Becca Roelofs, Anaïs White, Anders Andreassen, Tamara von Glehn, Lakshman Yagati, Mehran Kazemi, Lucas Gonzalez, Misha Khalman, Jakub Sygnowski, Alexandre Frechette, Charlotte Smith, Laura Culp, Lev Proleev, Yi Luan, Xi Chen, James Lottes, Nathan Schucher, Federico Lebron, Alban Rrustemi, Natalie Clay, Phil Crone, Tomas Kocisky, Jeffrey Zhao, Bartek Perz, Dian Yu, Heidi Howard, Adam Bloniarz, Jack W. Rae, Han Lu, Laurent Sifre, Marcello Maggioni, Fred Alcober, Dan Garrette, Megan Barnes, Shantanu Thakoor, Jacob Austin, Gabriel Barth-Maron, William Wong, Rishabh Joshi, Rahma Chaabouni, Deeni Fatiha, Arun Ahuja, Ruibo Liu, Yunxuan Li, Sarah Cogan, Jeremy Chen, Chao Jia, Chenjie Gu, Qiao Zhang, Jordan Grimstad, Ale Jakse Hartman, Martin Chadwick, Gaurav Singh Tomar, Xavier Garcia, Evan Senter, Emanuel Taropa, Thanumalayan Sankaranarayana Pillai, Jacob Devlin, Michael Laskin, Diego de Las Casas, Dasha Valter, Connie Tao, Lorenzo Blanco, Adrià Puigdomènech Badia, David Reitter, Mianna Chen, Jenny Brennan, Clara Rivera, Sergey Brin, Shariq Iqbal, Gabriela Surita, Jane Labanowski, Abhi Rao, Stephanie Winkler, Emilio Parisotto, Yiming Gu, Kate Olszewska, Yujing Zhang, Ravi Addanki, Antoine Miech, Annie Louis, Laurent El Shafey, Denis Teplyashin, Geoff Brown, Elliot Catt, Nithya Attaluri, Jan Balaguer, Jackie Xiang, Pidong Wang, Zoe Ashwood, Anton Briukhov, Albert Webson, Sanjay Ganapathy, Smit Sanghavi, Ajay Kannan, Ming-Wei Chang, Axel Stjerngren, Josip Djolonga, Yuting Sun, Ankur Bapna, Matthew Aitchison, Pedram Pejman, Henryk Michalewski, Tianhe Yu, Cindy Wang, Juliette Love, Junwhan Ahn, Dawn Bloxwich, Kehang Han, Peter Humphreys, Thibault Sellam, James Bradbury, Varun Godbole, Sina Samangooei, Bogdan Damoc, Alex Kaskasoli, Sébastien M. R. Arnold, Vijay Vasudevan, Shubham Agrawal, Jason Riesa, Dmitry Lepikhin, Richard Tanburn, Srivatsan Srinivasan, Hyeontaek Lim, Sarah Hodkinson, Pranav Shyam, Johan Ferret, Steven Hand, Ankush Garg, Tom Le Paine, Jian Li, Yujia Li, Minh Giang, Alexander Neitz, Zaheer Abbas, Sarah York, Machel Reid, Elizabeth Cole, Aakanksha Chowdhery, Dipanjan Das, Dominika Rogozińska, Vitaly Nikolaev, Pablo Sprechmann, Zachary Nado, Lukas Zilka, Flavien Prost, Luheng He, Marianne Monteiro, Gaurav Mishra, Chris Welty, Josh Newlan, Dawei Jia, Miltiadis Allamanis, Clara Huiyi Hu, Raoul de Liedekerke, Justin Gilmer, Carl Saroufim, Shruti Rijhwani, Shaobo Hou, Disha Shrivastava, Anirudh Baddepudi, Alex Goldin, Adnan Ozturel, Albin Cassirer, Yunhan Xu, Daniel Sohn, Devendra Sachan, Reinald Kim Amplayo, Craig Swanson, Dessie Petrova, Shashi Narayan, Arthur Guez, Siddhartha Brahma, Jessica Landon, Miteyan Patel, Ruizhe Zhao, Kevin Villela, Luyu Wang, Wenhao Jia, Matthew Rahtz, Mai Giménez, Legg Yeung, Hanzhao Lin, James Keeling, Petko Georgiev, Diana Mincu, Boxi Wu, Salem Haykal, Rachel Saputro, Kiran Vodrahalli, James Qin, Zeynep Cankara, Abhanshu Sharma, Nick Fernando, Will Hawkins, Behnam Neyshabur, Solomon Kim, Adrian Hutter, Priyanka Agrawal, Alex Castro-Ros, George van den Driessche, Tao Wang, Fan Yang, Shuo yiin Chang, Paul Komarek, Ross McIlroy, Mario Lučić, Guodong Zhang, Wael Farhan, Michael Sharman, Paul Natsev, Paul Michel, Yong Cheng, Yamini Bansal, Siyuan Qiao, Kris Cao, Siamak Shakeri, Christina Butterfield, Justin Chung, Paul Kishan Rubenstein, Shivani Agrawal, Arthur Mensch, Kedar Soparkar, Karel Lenc, Timothy Chung, Aedan Pope, Loren Maggiore, Jackie Kay, Priya Jhakra, Shibo Wang, Joshua Maynez, Mary Phuong, Taylor Tobin, Andrea Tacchetti, Maja Trebacz, Kevin Robinson, Yash Katariya, Sebastian Riedel, Paige Bailey, Kefan Xiao, Nimesh Ghelani, Lora Aroyo, Ambrose Slone, Neil Houlsby, Xuehan Xiong, Zhen Yang, Elena Gribovskaya, Jonas Adler, Mateo Wirth, Lisa Lee, Music Li, Thais Kagohara, Jay Pavagadhi, Sophie Bridgers, Anna Bortsova, Sanjay Ghemawat, Zafarali Ahmed, Tianqi Liu, Richard Powell, Vijay Bolina, Mariko Iinuma, Polina Zablotskaia, James Besley, Da-Woon Chung, Timothy Dozat, Ramona Comanescu, Xiance Si, Jeremy Greer, Guolong Su, Martin Polacek, Raphaël Lopez Kaufman, Simon Tokumine, Hexiang Hu, Elena Buchatskaya, Yingjie Miao, Mohamed Elhawaty, Aditya Siddhant, Nenad Tomasev, Jinwei Xing, Christina Greer, Helen Miller, Shereen Ashraf, Aurko Roy, Zizhao Zhang, Ada Ma, Angelos Filos, Milos Besta, Rory Blevins, Ted Klimenko, Chih-Kuan Yeh, Soravit Changpinyo, Jiaqi Mu, Oscar Chang, Mantas Pajarskas, Carrie Muir, Vered Cohen, Charline Le Lan, Krishna Haridasan, Amit Marathe, Steven Hansen, Sholto Douglas, Rajkumar Samuel, Mingqiu Wang, Sophia Austin, Chang Lan, Jiepu Jiang, Justin Chiu, Jaime Alonso Lorenzo, Lars Lowe Sjösund, Sébastien Cevey, Zach Gleicher, Thi Avrahami, Anudhyan Boral, Hansa Srinivasan, Vittorio Selo, Rhys May, Konstantinos Aisopos, Léonard Hussenot, Livio Baldini Soares, Kate Baumli, Michael B. Chang, Adrià Recasens, Ben Caine, Alexander Pritzel, Filip Pavetic, Fabio Pardo, Anita Gergely, Justin Frye, Vinay Ramasesh, Dan Horgan, Kartikeya Badola, Nora Kassner, Subhrajit Roy, Ethan Dyer, Víctor Campos, Alex Tomala, Yunhao Tang, Dalia El Badawy, Elspeth White, Basil Mustafa, Oran Lang, Abhishek Jindal, Sharad Vikram, Zhitao Gong, Sergi Caelles, Ross Hemsley, Gregory Thornton, Fangxiaoyu Feng, Wojciech Stokowiec, Ce Zheng, Phoebe Thacker, Çağlar Ünlü, Zhishuai Zhang, Mohammad Saleh, James Svensson, Max Bileschi, Piyush Patil, Ankesh Anand, Roman Ring, Katerina Tsihlas, Arpi Vezer, Marco Selvi, Toby Shevlane, Mikel Rodriguez, Tom Kwiatkowski, Samira Daruki, Keran Rong, Allan Dafoe, Nicholas FitzGerald, Keren Gu-Lemberg, Mina Khan, Lisa Anne Hendricks, Marie Pellat, Vladimir Feinberg, James Cobon-Kerr, Tara Sainath, Maribeth Rauh, Sayed Hadi Hashemi, Richard Ives, Yana Hasson, YaGuang Li, Eric Noland, Yuan Cao, Nathan Byrd, Le Hou, Qingze Wang, Thibault Sottiaux, Michela Paganini, Jean-Baptiste Lespiau, Alexandre Moufarek, Samer Hassan, Kaushik Shivakumar, Joost van Amersfoort, Amol Mandhane, Pratik Joshi, Anirudh Goyal, Matthew Tung, Andrew Brock, Hannah Sheahan, Vedant Misra, Cheng Li, Nemanja Rakićević, Mostafa Dehghani, Fangyu Liu, Sid Mittal, Junhyuk Oh, Seb Noury, Eren Sezener, Fantine Huot, Matthew Lamm, Nicola De Cao, Charlie Chen, Gamaleldin Elsayed, Ed Chi, Mahdis Mahdieh, Ian Tenney, Nan Hua, Ivan Petrychenko, Patrick Kane, Dylan Scandinaro, Rishub Jain, Jonathan Uesato, Romina Datta, Adam Sadovsky, Oskar Bunyan, Dominik Rabiej, Shimu Wu, John Zhang, Gautam Vasudevan, Edouard Leurent, Mahmoud Alnahlawi, Ionut Georgescu, Nan Wei, Ivy Zheng, Betty Chan, Pam G Rabinovitch, Piotr Stanczyk, Ye Zhang, David Steiner, Subhajit Naskar, Michael Azzam, Matthew Johnson, Adam Paszke, Chung-Cheng Chiu, Jaume Sanchez Elias, Afroz Mohiuddin, Faizan Muhammad, Jin Miao, Andrew Lee, Nino Vieillard, Sahitya Potluri, Jane Park, Elnaz Davoodi, Jiageng Zhang, Jeff Stanway, Drew Garmon, Abhijit Karmarkar, Zhe Dong, Jong Lee, Aviral Kumar, Luowei Zhou, Jonathan Evens, William Isaac, Zhe Chen, Johnson Jia, Anselm Levskaya, Zhenkai Zhu, Chris Gorgolewski, Peter Grabowski, Yu Mao, Alberto Magni, Kaisheng Yao, Javier Snaider, Norman Casagrande, Paul Suganthan, Evan Palmer, Geoffrey Irving, Edward Loper, Manaal Faruqui, Isha Arkatkar, Nanxin Chen, Izhak Shafran, Michael Fink, Alfonso Castaño, Irene Giannoumis, Wooyeol Kim, Mikołaj Rybiński, Ashwin Sreevatsa, Jennifer Prendki, David Soergel, Adrian Goedeckemeyer, Willi Gierke, Mohsen Jafari, Meenu Gaba, Jeremy Wiesner, Diana Gage Wright, Yawen Wei, Harsha Vashisht, Yana Kulizhskaya, Jay Hoover, Maigo Le, Lu Li, Chimezie Iwuanyanwu, Lu Liu, Kevin Ramirez, Andrey Khorlin, Albert Cui, Tian LIN, Marin Georgiev, Marcus Wu, Ricardo Aguilar, Keith Pallo, Abhishek Chakladar, Alena Repina, Xihui Wu, Tom van der Weide, Priya Ponnapalli, Caroline Kaplan, Jiri Simsa, Shuangfeng Li, Olivier Dousse, Fan Yang, Jeff Piper, Nathan Ie, Minnie Lui, Rama Pasumarthi, Nathan Lintz, Anitha Vijayakumar, Lam Nguyen Thiet, Daniel Andor, Pedro Valenzuela, Cosmin Paduraru, Daiyi Peng, Katherine Lee, Shuyuan Zhang, Somer Greene, Duc Dung Nguyen, Paula Kurylowicz, Sarmishta Velury, Sebastian Krause, Cassidy Hardin, Lucas Dixon, Lili Janzer, Kiam Choo, Ziqiang Feng, Biao Zhang, Achintya Singhal, Tejasi Latkar, Mingyang Zhang, Quoc Le, Elena Allica Abellan, Dayou Du, Dan McKinnon, Natasha Antropova, Tolga Bolukbasi, Orgad Keller, David Reid, Daniel Finchelstein, Maria Abi Raad, Remi Crocker, Peter Hawkins, Robert Dadashi, Colin Gaffney, Sid Lall, Ken Franko, Egor Filonov, Anna Bulanova, Rémi Leblond, Vikas Yadav, Shirley Chung, Harry Askham, Luis C. Cobo, Kelvin Xu, Felix Fischer, Jun Xu, Christina Sorokin, Chris Alberti, Chu-Cheng Lin, Colin Evans, Hao Zhou, Alek Dimitriev, Hannah Forbes, Dylan Banarse, Zora Tung, Jeremiah Liu, Mark Omernick, Colton Bishop, Chintu Kumar, Rachel Sterneck, Ryan Foley, Rohan Jain, Swaroop Mishra, Jiawei Xia, Taylor Bos, Geoffrey Cideron, Ehsan Amid, Francesco Piccinno, Xingyu Wang, Praseem Banzal, Petru Gurita, Hila Noga, Premal Shah, Daniel J. Mankowitz, Alex Polozov, Nate Kushman, Victoria Krakovna, Sasha Brown, MohammadHossein Bateni, Dennis Duan, Vlad Firoiu, Meghana Thotakuri, Tom Natan, Anhad Mohananey, Matthieu Geist, Sidharth Mudgal, Sertan Girgin, Hui Li, Jiayu Ye, Ofir Roval, Reiko Tojo, Michael Kwong, James Lee-Thorp, Christopher Yew, Quan Yuan, Sumit Bagri, Danila Sinopalnikov, Sabela Ramos, John Mellor, Abhishek Sharma, Aliaksei Severyn, Jonathan Lai, Kathy Wu, Heng-Tze Cheng, David Miller, Nicolas Sonnerat, Denis Vnukov, Rory Greig, Jennifer Beattie, Emily Caveness, Libin Bai, Julian Eisenschlos, Alex Korchemniy, Tomy Tsai, Mimi Jasarevic, Weize Kong, Phuong Dao, Zeyu Zheng, Frederick Liu, Fan Yang, Rui Zhu, Mark Geller, Tian Huey Teh, Jason Sanmiya, Evgeny Gladchenko, Nejc Trdin, Andrei Sozanschi, Daniel Toyama, Evan Rosen, Sasan Tavakkol, Linting Xue, Chen Elkind, Oliver Woodman, John Carpenter, George Papamakarios, Rupert Kemp, Sushant Kafle, Tanya Grunina, Rishika Sinha, Alice Talbert, Abhimanyu Goyal, Diane Wu, Denese Owusu-Afriyie, Cosmo Du, Chloe Thornton, Jordi Pont-Tuset, Pradyumna Narayana, Jing Li, Sabaer Fatehi, John Wieting, Omar Ajmeri, Benigno Uria, Tao Zhu, Yeongil Ko, Laura Knight, Amélie Héliou, Ning Niu, Shane Gu, Chenxi Pang, Dustin Tran, Yeqing Li, Nir Levine, Ariel Stolovich, Norbert Kalb, Rebeca Santamaria-Fernandez, Sonam Goenka, Wenny Yustalim, Robin Strudel, Ali Elqursh, Balaji Lakshminarayanan, Charlie Deck, Shyam Upadhyay, Hyo Lee, Mike Dusenberry, Zonglin Li, Xuezhi Wang, Kyle Levin, Raphael Hoffmann, Dan Holtmann-Rice, Olivier Bachem, Summer Yue, Sho Arora, Eric Malmi, Daniil Mirylenka, Qijun Tan, Christy Koh, Soheil Hassas Yeganeh, Siim Põder, Steven Zheng, Francesco Pongetti, Mukarram Tariq, Yanhua Sun, Lucian Ionita, Mojtaba Seyedhosseini, Pouya Tafti, Ragha Kotikalapudi, Zhiyu Liu, Anmol Gulati, Jasmine Liu, Xinyu Ye, Bart Chrzaszcz, Lily Wang, Nikhil Sethi, Tianrun Li, Ben Brown, Shreya Singh, Wei Fan, Aaron Parisi, Joe Stanton, Chenkai Kuang, Vinod Koverkathu, Christopher A. Choquette-Choo, Yunjie Li, TJ Lu, Abe Ittycheriah, Prakash Shroff, Pei Sun, Mani Varadarajan, Sanaz Bahargam, Rob Willoughby, David Gaddy, Ishita Dasgupta, Guillaume Desjardins, Marco Cornero, Brona Robenek, Bhavishya Mittal, Ben Albrecht, Ashish Shenoy, Fedor Moiseev, Henrik Jacobsson, Alireza Ghaffarkhah, Morgane Rivière, Alanna Walton, Clément Crepy, Alicia Parrish, Yuan Liu, Zongwei Zhou, Clement Farabet, Carey Radebaugh, Praveen Srinivasan, Claudia van der Salm, Andreas Fidjeland, Salvatore Scellato, Eri Latorre-Chimoto, Hanna Klimczak-Plucińska, David Bridson, Dario de Cesare, Tom Hudson, Piermaria Mendolicchio, Lexi Walker, Alex Morris, Ivo Penchev, Matthew Mauger, Alexey Guseynov, Alison Reid, Seth Odoom, Lucia Loher, Victor Cotruta, Madhavi Yenugula, Dominik Grewe, Anastasia Petrushkina, Tom Duerig, Antonio Sanchez, Steve Yadlowsky, Amy Shen, Amir Globerson, Adam Kurzrok, Lynette Webb, Sahil Dua, Dong Li, Preethi Lahoti, Surya Bhupatiraju, Dan Hurt, Haroon Qureshi, Ananth Agarwal, Tomer Shani, Matan Eyal, Anuj Khare, Shreyas Rammohan Belle, Lei Wang, Chetan Tekur, Mihir Sanjay Kale, Jinliang Wei, Ruoxin Sang, Brennan Saeta, Tyler Liechty, Yi Sun, Yao Zhao, Stephan Lee, Pandu Nayak, Doug Fritz, Manish Reddy Vuyyuru, John Aslanides, Nidhi Vyas, Martin Wicke, Xiao Ma, Taylan Bilal, Evgenii Eltyshev, Daniel Balle, Nina Martin, Hardie Cate, James Manyika, Keyvan Amiri, Yelin Kim, Xi Xiong, Kai Kang, Florian Luisier, Nilesh Tripuraneni, David Madras, Mandy Guo, Austin Waters, Oliver Wang, Joshua Ainslie, Jason Baldridge, Han Zhang, Garima Pruthi, Jakob Bauer, Feng Yang, Riham Mansour, Jason Gelman, Yang Xu, George Polovets, Ji Liu, Honglong Cai, Warren Chen, XiangHai Sheng, Emily Xue, Sherjil Ozair, Adams Yu, Christof Angermueller, Xiaowei Li, Weiren Wang, Julia Wiesinger, Emmanouil Koukoumidis, Yuan Tian, Anand Iyer, Madhu Gurumurthy, Mark Goldenson, Parashar Shah, MK Blake, Hongkun Yu, Anthony Urbanowicz, Jennimaria Palomaki, Chrisantha Fernando, Kevin Brooks, Ken Durden, Harsh Mehta, Nikola Momchev, Elahe Rahimtoroghi, Maria Georgaki, Amit Raul, Sebastian Ruder, Morgan Redshaw, Jinhyuk Lee, Komal Jalan, Dinghua Li, Ginger Perng, Blake Hechtman, Parker Schuh, Milad Nasr, Mia Chen, Kieran Milan, Vladimir Mikulik, Trevor Strohman, Juliana Franco, Tim Green, Demis Hassabis, Koray Kavukcuoglu, Jeffrey Dean, and Oriol Vinyals. 2023. Gemini: A family of highly capable multimodal models.

Anthropic. 2023. Claude 2.

Giorgio Bimbatti, Daniela Fogli, and Luigi Gargioni. 2023. Can chatgpt support end-user development of robot programs? *CEUR Workshop Proceedings*, 3408:2023.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the opportunities and risks of foundation models.

Petter Bae Brandtzaeg, Yukun You, Xi Wang, and Yucong Lao. 2023. "good" and "bad" machine agency in the context of human-ai communication: The case of chatgpt. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14059:3 – 23.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, Pete Florence, Chuyuan Fu, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Kehang Han, Karol Hausman, Alexander Herzog, Jasmine Hsu, Brian Ichter, Alex Irpan, Nikhil Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Isabel Leal, Lisa Lee, Tsang Wei Edward Lee, Sergey Levine, Yao Lu, Henryk Michalewski, Igor Mordatch, Karl Pertsch, Kanishka Rao, Krista Reymann, Michael Ryoo, Grecia Salazar, Pannag Sanketi, Pierre Sermanet, Jaspiar Singh, Anikait Singh, Radu Soricut, Huong Tran, Vincent Vanhoucke, Quan Vuong, Ayzaan Wahid, Stefan Welker, Paul Wohlhart, Jialin Wu, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Tianhe Yu, and Brianna Zitkovich. 2023. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *Proceedings of Machine Learning Research*, 229.

Edvard P.G. Bruun, Sigrid Adriaenssens, and Stefana Parascho. 2022. Structural rigidity theory applied to the scaffold-free (dis)assembly of space frames using cooperative robotics. *Automation in Construction*, 141:104405.

Yue Cao and C. S.George Lee. 2023. Robot behavior-tree-based task generation with large language models. *CEUR Workshop Proceedings*, 3433.

Yupeng Chang, X U Wang, Yuan Wu, Hao Chen, Xiaoyuan Yi, Wei Ye, Yue Zhang, Y I Chang, Philip S Yu, Yi Chang, Xu Wang, Jindong Wang, Linyi Yang, Kaijie Zhu, Cunxiang Wang, Yidong Wang, Qiang Yang, and Xing Xie. 2023. A survey on evaluation of large language models. *J. ACM*, 37:45.

Chaomei Chen and Min Song. 2019. Visualizing a field of research: A methodology of systematic scientometric reviews. *PLOS ONE*, 14:e0223994.

Young Min Cho, Sunny Rai, Lyle Ungar, João Sedoc, and Sharath Guntuku. 2023. An integrative survey on mental health conversational agents to bridge computer science and medical perspectives. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11346–11369, Singapore. Association for Computational Linguistics.

Yuchen Cui, Scott Niekum, Abhinav Gupta, Vikash Kumar, and Aravind Rajeswaran. 2022. Can foundation models perform zero-shot task specification for robot manipulation? *Proceedings of Machine Learning Research*, 168:893–905.

Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature 2015 521:7553*, 521:503–507.

Juan Manuel Davila Delgado and Lukumon Oyedele. 2022. Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms. *Advanced Engineering Informatics*, 54:101787.

Haolin Fan, Xuan Liu, Jerry Ying Hsi Fuh, Wen Feng Lu, and Bingbing Li. 2024. Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics. *Journal of Intelligent Manufacturing*.

Roya Firoozi, Johnathan Tucker, Stephen Tian, Anirudha Majumdar, Jiankai Sun, Weiyu Liu, Yuke Zhu, Shuran Song, Ashish Kapoor, Karol Hausman, Brian Ichter, Danny Driess, Jiajun Wu, Cewu Lu, and Mac Schwager. 2023. Foundation models in robotics: Applications, challenges, and the future.

Samir Yitzhak Gadre, Mitchell Wortsman, Gabriel Ilharco, Ludwig Schmidt, and Shuran Song. 2022. Cows on pasture: Baselines and benchmarks for language-driven zero-shot object navigation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023-June:23171–23181.

Michael Gusenbauer and Neal R. Haddaway. 2020. Which academic search systems are suitable for systematic reviews or meta-analyses? evaluating retrieval qualities of google scholar, pubmed, and 26 other resources. *Research Synthesis Methods*, 11:181–217.

Yafei Hu, Quanting Xie, Vidhi Jain, Jonathan Francis, Jay Patrikar, Nikhil Keetha, Seungchan Kim, Yaqi Xie, Tianyi Zhang, Shibo Zhao, Yu Quan Chong, Chen Wang, Katia Sycara, Matthew Johnson-Roberson, Dhruv Batra, Xiaolong Wang, Sebastian Scherer, Zsolt Kira, Fei Xia, Yonatan Bisk, Bosch AI Center for, Sair Lab, Georgia Tech, Fair at Meta, Uc San Diego, and Google DeepMind Robotics. 2023. Toward general-purpose robots via foundation models: A survey and meta-analysis robotic foundation models.

Raghav Jain, Daivik Sojitra, Arkadeep Acharya, Sriparna Saha, Adam Jatowt, and Sandipan Dandapat. 2023. Do language models have a common sense regarding time? revisiting temporal commonsense reasoning in the era of large language models. *EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 6750–6774.

Kris Jangjarat, Tanpat Kraiwanit, Pongsakorn Limna, and Rattaphong Sonsuphap. 2023. Public perceptions towards chatgpt as the robo-assistant. *Online Journal of Communication and Media Technologies*, 13:e202337.

Yixiang Jin, Dingzhe Li, A. V. Yong, Jun Shi, Peng Hao, Fuchun Sun, Jianwei Zhang, and Bin Fang. 2023. Robotgpt: Robot manipulation learning from chatgpt. *IEEE Robotics and Automation Letters*, 9:2543–2550.

Amita Kamath, Jack Hessel, and Kai Wei Chang. 2023. What's "up" with vision-language models? investigating their struggle with spatial reasoning. *EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 9161–9175.

S Kandall. 2017. Which one is best: Pubmed, web of science, or google scholar? - pubmed, web of science, or google scholar? a behind-the-scenes guide for life scientists. - libguides at michigan state university libraries.

Artur Istvan Karoly, Peter Galambos, Jozsef Kuti, and Imre J. Rudas. 2021. Deep learning in robotics: Survey on model structures and training strategies. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51:266–279.

Kento Kawaharazuka, Yoshiki Obinata, Naoaki Kanazawa, Kei Okada, and Masayuki Inaba. 2023. Robotic applications of pre-trained vision-language models to various recognition behaviors. *IEEE-RAS International Conference on Humanoid Robots*.

Stefanie Krause and Frieder Stolzenburg. 2024. Commonsense reasoning and explainable artificial intelligence using large language models. *Communications in Computer and Information Science*, 1947:302–319.

Yunshi Lan, Xiang Li, Xin Liu, Yang Li, Wei Qin, and Weining Qian. 2023. Improving zero-shot visual question answering via large language models with reasoning question prompts. *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pages 4389–4400.

Thomas Lew, Sumeet Singh, Mario Prats, Jeffrey Bingham, Jonathan Weisz, Benjie Holson, Xiaohan Zhang, Vikas Sindhwani, Yao Lu, Fei Xia, Peng Xu, Tingnan Zhang, Jie Tan, and Montserrat Gonzalez. 2023. Robotic table wiping via reinforcement learning and whole-body trajectory optimization. *Proceedings - IEEE International Conference on Robotics and Automation*, 2023-May:7184–7190.

Jianfeng Liao, Haoyang Zhang, Haofu Qian, Qiwei Meng, Yinan Sun, Yao Sun, Wei Song, Shiqiang Zhu, and Jason Gu. 2023. Decision-making in robotic grasping with large language models. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14271 LNAI:424–433.

Bingqian Lin, Yi Zhu, Zicong Chen, Xiwen Liang, Jianzhuang Liu, and Xiaodan Liang. 2022. Adapt: Vision-language navigation with modality-aligned action prompts. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June:15375–15385.

Yixin Liu, Kai Zhang, Yuan Li, Zhiling Yan, Chujie Gao, Ruoxi Chen, Zhengqing Yuan, Yue Huang, Hanchi Sun, Jianfeng Gao, Lifang He, and Lichao Sun. 2024. Sora: A review on background, technology, limitations, and opportunities of large vision models.

Chinmaya Mishra, Rinus Verdonschot, Peter Hagoort, and Gabriel Skantze. 2023. Real-time emotion generation in human-robot dialogue using large language models. *Frontiers in Robotics and AI*, 10.

Rajiv Movva, Sidhika Balachandar, Kenny Peng, Gabriel Agostini, Nikhil Garg, and Emma Pierson. 2023. Topics, authors, and institutions in large language model research: Trends from 17k arxiv papers.

Yoshiki Obinata, Kento Kawaharazuka, Naoaki Kanazawa, Naoya Yamaguchi, Naoto Tsukamoto, Iori Yanokura, Shingo Kitagawa, Koki Shinjo, Kei Okada, and Masayuki Inaba. 2023. Semantic scene difference detection in daily life patroling by mobile robots using pre-trained large-scale vision-language model. *IEEE International Conference on Intelligent Robots and Systems*, pages 3228–3233.

Felix Ocker, Jörg Deigmöller, and Julian Eggert. 2023. Exploring large language models as a source of common-sense knowledge for robots. *CEUR Workshop Proceedings*, 3632.

Siqi Ouyang and Lei Li. Autoplan: Automatic planning of interactive decision-making tasks with large language models.

Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10:1–11.

Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, Zhenjia Xu, Dorsa Sadigh, Andy Zeng, and Anirudha Majumdar. 2023. Robots that ask for help: Uncertainty alignment for large language model planners. *Proceedings of Machine Learning Research*, 229.

Jérôme Rutinowski, Sven Franke, Jan Endendyk, Ina Dormuth, Moritz Roidl, and Markus Pauly. 2023. The self-perception and political biases of chatgpt. *Human Behavior and Emerging Technologies*, 2024.

Kamel S. Saidi, Thomas Bock, and Christos Georgoulas. 2016. Robotics in construction. *Springer Handbooks*, pages 1493–1520.

Miran Seo, Samraat Gupta, and Youngjib Ham. 2024. Evaluation of work performance, task load, and behavior changes on time-delayed teleoperation tasks in space construction. *Construction Research Congress 2024, CRC 2024*, 1:89–98.

Javier Sevilla-Salcedo, Enrique Fernández-Rodicio, Laura Martín-Galván, Álvaro Castro-González, José C. Castillo, and Miguel A. Salichs. 2023. Using large language models to shape social robots' speech. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8:6–20.

Mohit Shridhar, Lucas Manuelli, and Dieter Fox. 2021. Cliport: What and where pathways for robotic manipulation. *Proceedings of Machine Learning Research*, 164:894–906.

Rishi Sinha. 2023. Statistical analysis of bias in chatgpt using prompt engineering. *International Journal for Research in Applied Science and Engineering Technology*, 11:1483–1489.

Tim Smithers. 1997. Autonomy in robots and other agents. *Brain and Cognition*, 34:88–106.

Austin Stone, Ted Xiao, Yao Lu, Keerthana Gopalakrishnan, Kuang Huei Lee, Quan Vuong, Paul Wohlhart, Sean Kirmani, Brianna Zitkovich, Fei Xia, Chelsea Finn, and Karol Hausman. 2023. Open-world object manipulation using pre-trained vision-language models. *Proceedings of Machine Learning Research*, 229.

Wilbert Tabone and Joost De Winter. 2023. Using chatgpt for human-computer interaction research: a primer. *Royal Society Open Science*, 10:231053.

Yuki Tanaka and Seiichiro Katsura. 2023. A voice-controlled motion reproduction using large language models for polishing robots. *Proceedings - 2023 IEEE International Conference on Mechatronics, ICM 2023*.

Yuan Tang, Xu Han, Xianzhi Li, Qiao Yu, Yixue Hao, Long Hu, and Min Chen. 2024. Minigpt-3d: Efficiently aligning 3d point clouds with large language models using 2d priors. 1.

J. M. Toro. 2023. Emergence of a phonological bias in chatgpt. *arXiv.org*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael, Smith Ranjan, Subramanian Xiaoqing, Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *ArXiv*.

Martijn Visser, Nees Jan van Eck, and Ludo Waltman. 2020. Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. *Quantitative Science Studies*, 2:20–41.

Xuan Xiao, Jiahang Liu, Zhipeng Wang, Yanmin Zhou, Yong Qi, Qian Cheng, Bin He, and Shuo Jiang. 2023. Robot learning in the era of foundation models: A survey.

Bing Xie, Xiangming Xi, Xinan Zhao, Yuhan Wang, Wei Song, Jianjun Gu, and Shiqiang Zhu. 2023. Chatgpt for robotics: A new approach to human-robot interaction and task planning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14271 LNAI:365–376.

Yang Ye, Hengxu You, and Jing Du. 2023. Improved trust in human-robot collaboration with chatgpt. *IEEE Access*, 11:55748–55754.

Naruki Yoshikawa, Marta Skreta, Kourosh Darvish, Sebastian Arellano-Rubach, Zhi Ji, Lasse Bjørn Kristensen, Andrew Zou Li, Yuchi Zhao, Haoping Xu, Artur Kuramshin, Alán Aspuru-Guzik, Florian Shkurti, and Animesh Garg. 2023. Large language models for chemistry robotics. *Autonomous Robots*, 47:1057–1086.

Bangguo Yu, Hamidreza Kasaei, and Ming Cao. 2023. L3mvn: Leveraging large language models for visual target navigation. *IEEE International Conference on Intelligent Robots and Systems*, pages 3554–3560.

Seung Nam Yu, Byung Gab Ryu, Sung Jin Lim, Chang Jun Kim, Maing Kyu Kang, and Chang Soo Han. 2009. Feasibility verification of brick-laying robot using manipulation trajectory and the laying pattern optimization. *Automation in Construction*, 18:644–655.

Kunpeng Zhang, Shipu Wang, Ning Jia, Liang Zhao, Chunyang Han, and Li Li. 2024. Integrating visual large language model and reasoning chain for driver behavior analysis and risk assessment. *Accident Analysis and Prevention*, 198.

Jingyan Zhou, Minda Hu, Junan Li, Xiaoying Zhang, Xixin Wu, Irwin King, and Helen Meng. 2023a. Rethinking machine ethics – can llms perform moral reasoning through the lens of moral theories?

Kankan Zhou, Eason Lai, Wei Bin Au Yeong, Kyriakos Mouratidis, and Jing Jiang. 2023b. Rome: Evaluating pre-trained vision-language models on reasoning beyond visual common sense. *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 10185–10197.