Leveraging Foundation Models in Robotics: Transforming Task Planning and Contextual Execution

Abiodun Sunday Adebayo [1*], Naomi Chukwurah [2], Olanrewaju Oluwaseun Ajayi [3]
[1] University of Staffordshire, United Kingdom
[2] Independent Researcher, USA
[3] University of the Cumberlands, USA

* Corresponding Author: **Abiodun Sunday Adebayo**

Article Info

ISSN (online): 2582-7138
Volume: 05
Issue: 01
January-February 2024
Received: 12-12-2023
Accepted: 08-01-2024
Page No: 1388-1394

Abstract
This paper explores integrating foundation models, particularly Transformers, into robotic systems to address challenges in task planning and contextual execution. Current robotic methodologies often struggle with adaptability and real-time decision-making in dynamic environments, limiting their effectiveness in complex tasks. Foundation models, known for their success in natural language processing and computer vision, offer significant potential to enhance robotic performance through improved contextual awareness and adaptability. The paper proposes a conceptual framework for incorporating these models into robotic architectures, detailing the necessary adaptations to model architecture, training techniques, and real-time sensor data integration. It also discusses expected outcomes, including increased precision, adaptability to new environments, and handling complex tasks. Finally, the paper identifies key areas for future research, such as exploring alternative foundation models, advancing training methodologies, and developing new evaluation metrics for robotic systems. This review underscores the transformative potential of foundation models in robotics and calls for continued innovation to realize their benefits fully.

DOI: https://doi.org/10.54660/.IJMRGE.2024.5.1.1388-1394

Keywords: Foundation Models, Robotics, Task Planning, Contextual Execution, Transformers

1. Introduction

1.1 Background and motivation

Robotics has seen remarkable advancements over the past few decades, with systems increasingly capable of performing complex tasks in diverse environments (Agarwala, 2020). Task planning and contextual execution are critical components of robotic functionality, enabling machines to make decisions, adapt to new information, and carry out tasks in dynamic settings. Effective task planning allows robots to sequence actions logically to achieve a specific goal. At the same time, contextual execution ensures that these actions are carried out in a manner that is appropriate to the surrounding environment. However, despite the progress in robotic technology, the field still faces significant challenges, particularly in achieving a high contextual awareness and adaptability level (Javaid, Haleem, Singh, & Suman, 2021).

Current approaches to task planning and contextual execution often rely on traditional machine-learning techniques and rule-based systems. While these methods have enabled robots to perform a wide range of tasks, they are often limited by their inability to generalize across different contexts or adapt to unforeseen circumstances. Traditional models typically require extensive hand-crafted features and domain-specific knowledge, making them inflexible and difficult to scale across different applications. Moreover, these models struggle to handle the complexity and variability inherent in real-world environments, leading to suboptimal performance in situations that deviate from the conditions under which they were trained (Sarker, Khan, Abushark, & Alsolami, 2021).

In contrast, foundation models, such as Transformers, have demonstrated remarkable capabilities in other fields, particularly in natural language processing (NLP) and computer vision. These models, characterized by their ability to process large amounts of data and learn generalizable representations, have revolutionized how tasks are approached in these domains. Transformers, for example, have become the backbone of state-of-the-art NLP systems, enabling machines to understand and generate human language with unprecedented accuracy. The success of these models in other areas raises the question of their potential applicability to robotics, particularly in enhancing task planning and contextual execution (Ekman, 2021; Tunstall, Von Werra, & Wolf, 2022). Despite the proven capabilities of foundation models in other domains, their adoption in robotics has been relatively limited. This gap presents a significant opportunity for innovation. The ability of foundation models to generalize across different tasks and contexts could address many of the limitations currently faced in robotic systems. By leveraging these models, it may be possible to develop robots that are more adaptable and capable of making more informed decisions in complex, dynamic environments. This paper seeks to explore this potential by proposing a conceptual framework for integrating foundation models into robotic systems to transform task planning and contextual execution.

1.2 Objectives

The primary objective of this paper is to bridge the gap between the capabilities of foundation models and their application in robotics. Specifically, the paper aims to propose a conceptual framework that integrates these models into robotic systems to enhance both task planning and contextual execution. The framework will draw on the successes of foundation models in other domains, such as NLP and computer vision, adapting these approaches to the unique challenges of robotics.

To achieve this objective, the paper will begin by reviewing current applications of foundation models in various fields, highlighting their key strengths and the reasons for their success. This review will provide a foundation for understanding how these models can be adapted to robotics. The paper will then explore the specific adaptations required to integrate foundation models into robotic systems, considering factors such as model architecture, training processes, and the incorporation of sensor data for real-time decision-making.

In addition to proposing a framework, the paper will outline future research directions that could further enhance the integration of foundation models into robotics. This will include suggestions for improving model training techniques, exploring new types of foundation models, and developing more sophisticated evaluation metrics to assess the performance of robotic systems. The ultimate goal is to pave the way for more adaptable, precise, and context-aware robotic systems that operate effectively in various environments.

By addressing these objectives, this paper aims to contribute to the growing body of research on applying advanced machine-learning techniques in robotics. It seeks to demonstrate that foundation models, which have already proven transformative in other fields, have the potential to enhance the capabilities of robotic systems significantly. In doing so, it hopes to inspire further research and development in this area, ultimately leading to more intelligent and adaptable robots that can better meet the demands of real-world applications.

The proposed framework is not intended to be a one-size-fits-all solution but rather a starting point for exploring the potential of foundation models in robotics. As such, the paper will emphasize the importance of ongoing research and experimentation to refine and optimize these models for specific robotic tasks. By building on the foundation laid by this paper, researchers and practitioners in the field of robotics can continue to push the boundaries of what is possible, ultimately leading to more capable and versatile robotic systems.

2. Foundation Models: An overview

2.1 Definition and core concepts

Foundation models represent a significant advancement in artificial intelligence (AI) and machine learning (ML), characterized by their ability to generalize across various tasks and domains. Unlike traditional machine learning models, which are often task-specific and require extensive domain knowledge for effective deployment, foundation models are designed to be versatile, capable of learning from vast amounts of data and applying this knowledge to various contexts. The term "foundation" reflects their role as a foundational layer upon which various applications can be built, leveraging the model's broad understanding to tackle diverse problems (Akrout, Feriani, Bellili, Mezghani, & Hossain, 2023; Bommasani *et al*., 2021).

Their architecture is at the heart of foundation models, which typically employ deep learning techniques to construct models with millions or even billions of parameters. One of the most notable examples of a foundation model is the Transformer, a model initially introduced for natural language processing tasks but has since been adapted for various other domains. The Transformer architecture is based on self-attention mechanisms, allowing the model to weigh the importance of different words in a sentence relative to each other. This capability enables the Transformer to understand context and relationships between elements in a sequence, making it particularly powerful for tasks that require contextual understanding (Tunstall *et al*., 2022).

Transformers consist of an encoder-decoder structure, where the encoder processes input data (such as a sentence), and the decoder generates the output (such as a translated sentence). The self-attention mechanism within the Transformer allows it to consider the entire input sequence when making predictions rather than just focusing on the immediate vicinity of a given word or token. This global attention mechanism is a key factor in the model's ability to capture long-range dependencies and relationships, a significant advantage over traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which are more limited in their capacity to handle such dependencies (Khan *et al*., 2022; Shreyashree, Sunagar, Rajarajeswari, & Kanavalli, 2022; Tunstall *et al*., 2022).

Another critical feature of foundation models, particularly Transformers, is their scalability. The architecture is designed to handle large datasets. It can be scaled up to accommodate more data and parameters, resulting in models with increasing levels of performance and generalization. This scalability has led to the developing of models like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers), which have set new benchmarks in NLP tasks. These models are pre-trained on massive datasets and can be fine-tuned for specific tasks, making them highly adaptable and efficient (Patwardhan, Marrone, & Sansone, 2023; Raiaan *et al*., 2024).

The difference between foundation and traditional machine learning models lies in their approach to learning and generalization. Traditional models are often trained on specific datasets for specific tasks, with performance highly dependent on the quality and relevance of the training data. These models typically require extensive feature engineering and domain expertise to perform well. In contrast, foundation models are designed to learn broadly applicable representations from large, diverse datasets. This pre-training allows them to generalize across various tasks with minimal fine-tuning, making them more robust and versatile in handling different types of data and challenges.

2.2 Applications in various domains

The versatility and adaptability of foundation models have made them highly successful in numerous fields beyond their initial application in NLP. One of the most prominent areas where foundation models have had a transformative impact is computer vision. In this domain, models like Vision Transformers (ViTs) have emerged, applying the principles of Transformers to image-processing tasks. ViTs treat images as sequences of patches analogous to words in a sentence, allowing the model to learn the spatial relationships between different parts of an image. This approach has led to significant improvements in tasks such as image classification, object detection, and image segmentation, where understanding the context and relationships within an image is crucial (K. Han *et al*., 2022; X. Han *et al*., 2021; Khan *et al*., 2022).

In autonomous systems, foundation models have also demonstrated considerable potential. Autonomous vehicles, for instance, require sophisticated perception and decision-making capabilities to navigate complex environments safely. Foundation models, with their ability to process and integrate information from multiple sensors (such as cameras, LiDAR, and radar), offer a way to enhance these vehicles' situational awareness and decision-making processes. By learning from large datasets encompassing various driving scenarios, these models can be generalized to new and unseen situations, improving the reliability and safety of autonomous driving systems (Ignatious, El-Sayed, Khan, & Mokhtar, 2023).

Moreover, foundation models have been leveraged in healthcare to enhance diagnostic accuracy and treatment planning. Models like GPT-3, while primarily designed for text generation, have been adapted to generate medical reports, assist in analyzing patient data, and even suggest possible diagnoses based on symptoms. These applications demonstrate the model's ability to synthesize vast amounts of information and provide contextually relevant insights, a capability that is invaluable in complex and high-stakes environments like healthcare (Nazi & Peng, 2024; Vavekanand, Karttunen, Xu, Milani, & Li, 2024).

In the domain of robotics, while foundation models have not yet been widely adopted, their potential is significant. Robotics requires the integration of perception, decision-making, and action, often in real-time and under varying conditions. Foundation models, with their ability to generalize across tasks and adapt to new contexts, are well-suited to address the challenges of robotic systems. For example, foundation models could be used in robotic vision to enhance object recognition and scene understanding, enabling robots to navigate and interact with their environments more effectively. In task planning, these models could be applied to develop more flexible and adaptive strategies that can handle the unpredictability and complexity of real-world tasks (Mota, Sridharan, & Leonardis, 2021).

The adaptability of foundation models also makes them promising candidates for enhancing contextual execution in robotics. Contextual execution involves understanding and responding to the nuances of the environment in which a task is performed. Foundation models, with their ability to capture and utilize context from large datasets, could enable robots to understand the specific circumstances of a task better and adjust their actions accordingly. This capability could significantly improve the precision and reliability of robotic systems, particularly in dynamic or uncertain environments.

Overall, the success of foundation models in various domains underscores their potential for broader application, including in robotics. By leveraging their ability to learn from diverse datasets, generalize across tasks, and adapt to new contexts, foundation models offer a promising path forward for enhancing the capabilities of robotic systems. As research in this area continues to evolve, these models will likely play an increasingly important role in advancing the field of robotics, leading to more intelligent, adaptable, and context-aware robots that can operate effectively in various environments.

3. Challenges and opportunities in robotics

3.1 Current state of task planning and contextual execution

Robotic systems have made significant strides in recent years, particularly in their ability to perform complex tasks across various environments. Task planning and contextual execution are at the core of these advancements, enabling robots to sequence actions, adapt to new information, and execute tasks in dynamic settings. However, despite these developments, existing methods for task planning and contextual execution are not without their limitations, and the field continues to grapple with challenges related to adaptability, precision, and contextual awareness (Mikolajczyk *et al*., 2022).

Traditional approaches to task planning in robotics often rely on rule-based systems and classical artificial intelligence techniques, such as search algorithms and finite state machines. These methods generate a sequence of actions a robot must follow to achieve a specific goal. While effective in structured environments with predictable conditions, these approaches struggle in more complex and dynamic settings. The rigidity of rule-based systems means that robots may fail to adapt when encountering unexpected obstacles or environmental changes, leading to suboptimal or even failed task execution (Javaid *et al*., 2021).

Contextual execution, on the other hand, involves the robot's ability to interpret and respond to the nuances of its environment during task performance. This requires a deep understanding of the context in which a task is carried out, including the relationships between different objects, the intentions behind human actions, and the potential consequences of different courses of action. Current methods for achieving contextual execution often involve machine learning models trained on specific datasets to recognize patterns and make predictions. However, these models are typically limited by their dependence on the quality and diversity of the training data. In environments that differ significantly from the training scenarios, the performance of these models can degrade, leading to incorrect decisions and actions (Lakshmanan, Robinson, & Munn, 2020; Xi & Zhu, 2023).

The complexity of contextual awareness further exacerbates these challenges. A robot operating in a real-world environment must contend with many variables, including lighting changes, obstacles, object appearance variations, and

the unpredictable behavior of humans and other robots. The sheer diversity of possible scenarios makes it difficult to create models that can accurately predict and respond to every potential situation. As a result, robots often lack the flexibility and adaptability needed to operate effectively in unstructured environments, limiting their utility in many real-world applications (Neu, Lahann, & Fettke, 2022). Moreover, integrating task planning and contextual execution remains an ongoing challenge. These two components are often developed and implemented separately, leading to disjointed systems where planning and execution are not fully aligned. This separation can result in inefficient task performance, as the robot may plan actions that are not well-suited to the specific context in which they are executed. The lack of a unified approach to task planning and contextual execution highlights the need for more advanced models to integrate these processes seamlessly, allowing robots to plan and execute tasks contextually appropriately and adapt to changes in their environment (Cacace, Caccavale, Finzi, & Grieco, 2023).

3.2 Potential of foundation models in robotics

The limitations of current approaches to task planning and contextual execution in robotics underscore the potential for innovation by integrating foundation models. Foundation models, such as Transformers, offer a promising solution to challenges faced by traditional methods. These models are characterized by their ability to generalize across a wide range of tasks and contexts, making them well-suited to the demands of robotic systems operating in dynamic and unpredictable environments.

One of the key theoretical advantages of foundation models is their capacity for improved adaptability. Unlike traditional models, which are often rigid and task-specific, foundation models are designed to learn from vast amounts of data, allowing them to capture a broad range of patterns and relationships. This generalization capability enables foundation models to adapt more easily to new and unforeseen situations. For example, a Transformer-based model could be trained on a diverse dataset that includes various environmental conditions, object types, and human behaviors. This training would allow the model to develop a deep understanding of the contextual factors that influence task performance, enabling it to adapt its actions in real-time based on the specific circumstances of the task (Yang *et al*., 2023).

In addition to adaptability, foundation models offer significant potential for enhancing real-time decision-making in robotics. The architecture of models like Transformers allows for parallel information processing, enabling them to analyze and integrate data from multiple sources quickly. This capability is particularly valuable in robotics, where timely decision-making is critical to successful task execution. By leveraging the processing power of foundation models, robotic systems can make faster and more informed decisions, reducing the likelihood of errors and improving overall performance (Bommasani *et al*., 2021).

Another important advantage of foundation models is their ability to enhance contextual understanding. As discussed earlier, contextual execution requires a robot to interpret the nuances of its environment and adjust its actions accordingly. Foundation models excel in this area because they can learn complex relationships between different elements within a dataset. For example, a Transformer-based model could analyze the spatial relationships between objects in a scene, the intent behind human actions, or the potential consequences of different decisions. This deep contextual understanding allows the model to generate more accurate and contextually appropriate responses, improving the robot's ability to execute tasks in a manner that is both effective and safe (Dennler *et al*., 2023; Xi & Zhu, 2023). Moreover, the scalability of foundation models presents an opportunity for developing more robust and versatile robotic systems. As these models are trained on increasingly large and diverse datasets, their ability to generalize and adapt to new tasks continues to improve. This scalability means foundation models can be continually updated and refined as new data becomes available, ensuring that robotic systems can handle the ever-evolving challenges of real-world environments (Kawaharazuka *et al*., 2024).

Integrating foundation models into robotic systems also offers the potential for a more unified approach to task planning and contextual execution. By leveraging the same model for planning and execution, creating systems where these processes are more closely aligned becomes possible. For example, a foundation model could generate a task plan based on an initial analysis of the environment and then continuously update the plan as the task is executed, considering new information and changes in the environment. This integrated approach would allow for more efficient and contextually appropriate task performance, addressing one of the key limitations of current robotic systems.

4. Proposed conceptual framework

4.1 Framework Design

Integrating foundation models into robotic systems presents an innovative approach to overcoming the limitations of current task planning and contextual execution methodologies. The proposed conceptual framework envisions a robotic architecture that leverages the strengths of foundation models, particularly their capacity for generalization and contextual understanding, to enhance the robot's ability to perform tasks in dynamic and complex environments.

The framework is built around several key components that enable more sophisticated task planning and execution. The foundation model, such as a Transformer-based architecture, is at the framework's core, which serves as the primary decision-making engine. This model is responsible for processing input data, generating task plans, and adjusting execution strategies based on contextual information in real-time.

The first component of the framework is the data acquisition and preprocessing module, which collects and processes data from various sensors, including cameras, LiDAR, radar, and other environmental sensors. This module ensures that the foundation model receives high-quality, context-rich data that accurately represents the robot's operating environment. The preprocessing steps may involve noise reduction, normalization, and converting raw sensor data into formats the model can efficiently process, such as sequences or vectors.

Next is the contextual analysis and understanding module, which employs the foundation model to analyze the processed data and extract meaningful contextual information. This module leverages the self-attention mechanisms of Transformers to understand relationships between different elements in the data, such as the spatial positioning of objects, the sequence of events in a task, and the intentions of human actors in the environment. The ability to process this information in parallel and at multiple levels of abstraction allows the model to understand the context in which it operates comprehensively.

The third component is the task planning module, which uses the insights gained from the contextual analysis to generate a sequence of actions the robot should follow to achieve its goals. Unlike traditional task planning systems, which often rely on predefined rules or heuristics, this module uses the foundation model's understanding of context to create flexible and adaptable plans. The task planning module continuously updates its plan as new data becomes available, ensuring that the robot's actions remain aligned with the current context.

Finally, the execution and feedback module is responsible for implementing the planned actions and monitoring their outcomes. This module integrates with the robot's actuators and control systems, translating high-level plans into specific commands that drive the robot's movements and interactions. As the robot executes its tasks, the module provides real-time feedback to the foundation model, allowing it to adjust the plan as necessary. This feedback loop is crucial for handling unexpected events and ensuring the robot can adapt to unforeseen challenges without manual intervention.

4.2 Adaptation of models for robotics

While foundation models like Transformers have proven their effectiveness in fields such as natural language processing and computer vision, their direct application to robotics requires certain adaptations to meet the unique demands of robotic systems. One of the primary considerations is modifying model architecture to accommodate the diverse and often continuous nature of sensor data in robotics. Unlike text or images, sensor data in robotics can vary widely in format, frequency, and scale, necessitating adjustments to how the model processes and integrates this information.

One potential adaptation is incorporating multi-modal processing capabilities, enabling the foundation model to simultaneously handle inputs from multiple sensors. This could involve extending the model's architecture to include specialized layers or modules for processing different data types, such as visual data from cameras, depth data from LiDAR, and positional data from GPS. These layers would fuse the sensor data into a unified representation that the model can use for task planning and contextual understanding.

Another crucial adaptation involves the training process. Traditional foundation models are often trained on large, static datasets collected and labeled in advance. However, robotic systems operate in dynamic environments where data constantly changes. To address this, the training process for foundation models in robotics may need to incorporate elements of online learning, where the model continues to learn and update its parameters as it interacts with the environment. This approach allows the model to remain flexible and responsive to new data, ensuring it can adapt to real-time changes.

The integration of sensor data for real-time contextual awareness is another critical adaptation. In robotics, responding to environmental changes on the fly is essential for effective task execution. Foundation models in this framework must be able to process sensor data in real time, making rapid decisions based on the latest contextual information. This might involve optimizing the model's inference processes to reduce latency and increase computational efficiency, allowing the robot to react quickly to dynamic situations.

Additionally, the foundation model may require specialized pre-training to develop a strong understanding of the specific contexts and tasks relevant to robotics. This could involve training the model on synthetic datasets generated by simulations replicating real-world robotic environments. These simulations can expose the model to various scenarios, including rare or extreme conditions that would be difficult to capture in real-world data. The model can develop robust generalization capabilities by pre-training on such diverse scenarios that enhance its performance in real-world applications.

4.3 Expected Outcomes

The integration of foundation models into robotic systems, as outlined in this conceptual framework, is expected to yield several significant improvements in task planning and contextual execution. One of the most notable outcomes is increased precision in task execution. By leveraging the advanced contextual understanding provided by foundation models, robots can make more informed decisions that are closely aligned with the specific requirements of each task, reducing the likelihood of errors and enhancing overall performance.

Another expected outcome is greater adaptability to new environments. Foundation models, with their ability to generalize across diverse contexts, are well-suited to handling the variability and unpredictability of real-world environments. This adaptability means that robots equipped with foundation models can operate effectively in new or unfamiliar settings without extensive reprogramming or manual intervention. As a result, these robots can be deployed in a wider range of applications, from manufacturing and logistics to healthcare and service industries.

The framework also anticipates improved handling of complex tasks. Tasks that involve multiple steps, interactions with humans, or coordination between different robots can be challenging for traditional systems to manage effectively. However, the ability of foundation models to understand and process complex relationships between different elements of a task allows for more sophisticated planning and execution strategies. This capability is particularly valuable in scenarios where the robot must respond to changes or interruptions, ensuring it can complete the task successfully even in the face of unforeseen challenges.

Finally, integrating foundation models will result in more efficient and autonomous robotic systems. By enabling robots to plan and execute tasks with minimal human oversight, foundation models can reduce the need for constant supervision and manual control. This autonomy improves the efficiency of robotic operations. It frees human operators to focus on tasks like system monitoring and decision-making. Over time, this could lead to a shift in the role of humans in robotic systems from direct control to strategic oversight, further enhancing the capabilities and effectiveness of robotic technologies.

5. Future research directions and conclusion

5.1 Research Directions

Integrating foundation models into robotic systems marks a significant advancement in robotics. Nevertheless, there remains ample scope for future research to enhance this integration further. One promising area of research involves exploring different foundation models beyond the commonly used Transformers. For instance, models such as BERT or GPT could be adapted and evaluated for their effectiveness in various robotic applications, potentially offering unique advantages in specific contexts or tasks. Investigating the suitability of these models in robotics could lead to the development of more specialized and efficient architectures tailored to the unique demands of robotic systems.

Another critical area for future research is the improvement of model training techniques. While current training methodologies have been successful in many domains, the dynamic and unpredictable nature of real-world environments presents unique challenges for robotics. Research could focus on developing hybrid training approaches combining offline pre-training with online learning, enabling robots to adapt to new data and scenarios continuously. Additionally, creating more sophisticated simulation environments for training foundation models could help expose them to a wider range of potential scenarios, thereby improving their generalization capabilities when deployed in real-world settings.

Developing new evaluation metrics for robotic performance is another crucial research direction. Traditional metrics used in other domains, such as accuracy or loss functions, may not fully capture the complexities of robotic task execution and contextual understanding. Future research could focus on designing metrics that better assess a robot's ability to perform tasks in dynamic environments, considering adaptability, real-time decision-making, and the ability to learn from experience. These metrics could provide more meaningful insights into the effectiveness of foundation models in robotics and guide the development of more advanced and capable systems.

Moreover, the research could explore collaborative learning techniques where multiple robots equipped with foundation models can share knowledge and experiences. This would enable the collective learning of models across different environments and tasks, accelerating the adaptation process and improving overall system robustness. Such approaches could be particularly valuable in industrial settings, where robots often work together to perform complex tasks.

5.2 Conclusion

Integrating foundation models into robotic systems represents a transformative leap forward in the capabilities of robots, particularly in task planning and contextual execution. Throughout this paper, we have explored the potential of foundation models like Transformers to enhance robotic systems' precision, adaptability, and contextual awareness, addressing many of the limitations inherent in traditional approaches.

The proposed conceptual framework outlines how these models can be effectively integrated into robotic architectures, with key components such as data acquisition, contextual analysis, task planning, and execution working harmoniously to create more intelligent and autonomous systems. By adapting foundation models to the unique demands of robotics, including real-time processing and multi-modal data integration, we can unlock new levels of performance and versatility in robotic applications.

However, the journey towards fully realizing the potential of foundation models in robotics is still ongoing. Future research in areas such as model exploration, training technique enhancements, and the development of new evaluation metrics will be crucial in pushing the boundaries of what robotic systems can achieve. By continuing to innovate and refine these models, the field of robotics benefits immensely, leading to robots that can perform complex tasks with a level of precision and contextual understanding that was previously unattainable.

6. References

1. Agarwala N. Monitoring the ocean environment using robotic systems: Advancements, trends, and challenges. Marine Technology Society Journal. 2020;54(5):42–60.
2. Akrout M, Feriani A, Bellili F, Mezghani A, Hossain E. Domain generalization in machine learning models for wireless communications: Concepts, state-of-the-art, and open issues. IEEE Communications Surveys & Tutorials. 2023.
3. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, *et al*. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. 2021.
4. Cacace J, Caccavale R, Finzi A, Grieco R. Combining human guidance and structured task execution during physical human–robot collaboration. Journal of Intelligent Manufacturing. 2023;34(7):3053–67.
5. Dennler N, Ruan C, Hadiwijoyo J, Chen B, Nikolaidis S, Matarić M. Design metaphors for understanding user expectations of socially interactive robot embodiments. ACM Transactions on Human-Robot Interaction. 2023;12(2):1–41.
6. Ekman M. Learning deep learning: Theory and practice of neural networks, computer vision, natural language processing, and transformers using TensorFlow. Addison-Wesley Professional; 2021.
7. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, *et al*. A survey on vision transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022;45(1):87–110.
8. Han X, Zhang Z, Ding N, Gu Y, Liu X, Huo Y, *et al*. Pre-trained models: Past, present and future. AI Open. 2021;2:225–50.
9. Ignatious HA, El-Sayed H, Khan MA, Mokhtar BM. Analyzing factors influencing situation awareness in autonomous vehicles—A survey. Sensors. 2023;23(8):4075.
10. Javaid M, Haleem A, Singh RP, Suman R. Substantial capabilities of robotics in enhancing Industry 4.0 implementation. Cognitive Robotics. 2021;1:58–75.
11. Kawaharazuka K, Matsushima T, Gambardella A, Guo J, Paxton C, Zeng A. Real-world robot applications of foundation models: A review. arXiv preprint arXiv:2402.05741. 2024.
12. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: A survey. ACM Computing Surveys (CSUR). 2022;54(10s):1–41.
13. Lakshmanan V, Robinson S, Munn M. Machine learning design patterns. O'Reilly Media; 2020.
14. Mikolajczyk T, Mikołajewska E, Al-Shuka HF, Malinowski T, Kłodowski A, Pimenov DY, *et al*. Recent advances in bipedal walking robots: Review of gait, drive, sensors and control systems. Sensors. 2022;22(12):4440.
15. Mota T, Sridharan M, Leonardis A. Integrated commonsense reasoning and deep learning for transparent decision making in robotics. SN Computer Science. 2021;2(4):242.
16. Nazi ZA, Peng W. Large language models in healthcare and medical domain: A review. Paper presented at: Informatics; 2024.
17. Neu DA, Lahann J, Fettke P. A systematic literature review on state-of-the-art deep learning methods for process prediction. Artificial Intelligence Review. 2022;55(2):801–27.
18. Patwardhan N, Marrone S, Sansone C. Transformers in the real world: A survey on NLP applications. Information. 2023;14(4):242.
19. Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, *et al*. A review on large language models: Architectures, applications, taxonomies, open issues and

challenges. IEEE Access. 2024.
20. Sarker IH, Khan AI, Abushark YB, Alsolami F. Mobile expert system: Exploring context-aware machine learning rules for personalized decision-making in mobile applications. Symmetry. 2021;13(10):1975.
21. Shreyashree S, Sunagar P, Rajarajeswari S, Kanavalli A. A literature review on bidirectional encoder representations from transformers. Inventive Computation and Information Technologies: Proceedings of ICICIT 2021. 2022;305–20.
22. Tunstall L, Von Werra L, Wolf T. Natural language processing with transformers. O'Reilly Media, Inc.; 2022.
23. Vavekanand R, Karttunen P, Xu Y, Milani S, Li H. Large language models in healthcare decision support: A review. 2024.
24. Xi X, Zhu S. A comprehensive review of task understanding of command-triggered execution of tasks for service robots. Artificial Intelligence Review. 2023;56(7):7137–93.
25. Yang M, Wu C, Guo Y, Jiang R, Zhou F, Zhang J, *et al*. Transformer-based deep learning model and video dataset for unsafe action identification in construction projects. Automation in Construction. 2023;146:104703.