A Comprehensive Survey on Embodied Artificial Intelligence: Foundations, Advances, and Future Directions

SurveyForge

Abstract— Embodied Artificial Intelligence (Embodied AI) represents a pivotal juncture in AI research, integrating cognitive and physical processes to enhance situational awareness and autonomy. This survey comprehensively examines Embodied AI's transformative evolution, addressing core dimensions such as sensorimotor integration, multi-modal perception, and adaptive learning. These systems challenge traditional AI by embodying cognition within physical realities, fostering enhanced interaction capabilities and learning efficiency. Key findings highlight advancements in co-optimizing morphology and control, reflecting the importance of aligning physical and cognitive elements for robust performance. However, challenges remain, particularly in sim-to-real transfer, where gaps between simulation and real-world performance need bridging through innovative domain randomization techniques and physics-informed models. The implications of Embodied AI extend to various application areas, including autonomous systems, assistive robotics, and human-robot collaboration. Future research will benefit from interdisciplinary efforts, leveraging insights from neuroscience and cognitive science to perfect embodied systems' design, ultimately paving the way toward Artificial General Intelligence (AGI). By aligning technical advancements with practical applications, Embodied AI sets the stage for a profound impact on human-machine interaction, offering a robust framework for intelligent behavior in dynamic environments.

Index Terms—Embodied AI systems, sensorimotor integration, adaptive learning mechanisms

1 INTRODUCTION

EMBODIED Artificial Intelligence (EAI) represents a transformative shift in the AI paradigm, moving beyond traditional cognition-focused approaches to systems that integrate cognitive processes with physical capabilities. This subsection provides a comprehensive introduction to EAI, examining its evolution, core principles, and its bridging role between cognitive and physical interactions. By incorporating insights from multiple disciplines, EAI has catalyzed a new wave of AI research that considers not just the mechanics of computation, but also the crucial interplay between an agent's body, its environment, and its decision-making processes.

The concept of embodiment in AI draws on the foundational idea that intelligence must be situated within a body to interact effectively with the world, aligning it closely with principles derived from cognitive science and robotics. This leads us to a critical assessment of the evolutionary trajectory of EAI. Historically, AI research primarily focused on symbolic manipulation and abstract problem-solving, disconnected from real-world applications. However, the inadequacies of this approach in addressing complex, dynamic environments necessitated a shift toward systems that can physically interact with their surroundings, thereby acknowledging the significance of physical embodiment in intelligent behavior. Early milestones in the evolution of EAI include biological-inspired models that emphasized the role of the body's morphology in cognition, as elucidated in [1].

Contemporary EAI frameworks seek to integrate both the morphological and cognitive paradigms by leveraging multi-modal sensing and actuation capabilities. Key principles of EAI involve understanding how sensorimotor interactions contribute to intelligent behavior. Sensorimotor loops, which encompass feedback mechanisms between sensory inputs and motor actions, are pivotal in shaping an agent's adaptive and learning behaviors [2]. The dynamic interaction between an agent's physical embodiment and its environment leads to emergent behaviors that often surpass computational models based solely on abstract cognition.

Academic analysis of EAI highlights several contrasting approaches within the field. On one hand, embodied cognition posits that intelligence emerges from the interactions between an organism's body and its environment, emphasizing the role of 'body schema' and 'forward models' in perception and action planning [3]. On the other hand, traditional AI approaches typically isolate cognition from physical embodiment, focusing on discrete learning tasks that can be solved computationally. The emergence of frameworks like Symbol Emergence in Robotics (SER) further exemplifies the integration of physical interaction and language learning within socially mediated contexts, pushing the boundaries of traditional cognition models [4].

A major strength of the embodied AI approach is its ability to facilitate naturalistic interaction in dynamic environments, leveraging the agent's physical form to reduce computational demands, a notion known as "cheap design" [5]. However, a notable limitation includes the high complexity in co-optimizing an agent's body and control systems, a challenge that is inextricably linked to trade-offs in adaptive behavior and morphological design [6]. Despite these challenges, significant advancements in simulation platforms and embodied cognitive architectures have mitigated some of these limitations, enhancing the feasibility of

EAI applications in real-world scenarios [7].

Emerging trends in EAI research underscore the growing importance of integrating causality and multi-sensory input processing to enhance agent adaptability and robustness in unforeseen environments. By embracing a more holistic view of intelligence that incorporates both the physical and cognitive aspects, EAI promises to significantly contribute to the development of Artificial General Intelligence (AGI) [8].

To synthesize, Embodied AI serves as a vital bridge between cognitive processes and physical realities, facilitated by advancements in sensor integration, computational models, and the evolving understanding of embodiment. As illustrated in numerous case studies and theoretical frameworks, the path forward for EAI involves continued interdisciplinary collaboration and innovation, particularly in enhancing co-evolutionary paradigms and simulation-to-reality transfer methods [9], [10]. Future research directions may focus on expanding the application of EAI in areas such as autonomous vehicles, assistive robotics, and human-robot interaction, leveraging the full potential of embodied systems to navigate complex societal and environmental challenges. In conclusion, the comprehensive survey and exploration of EAI not only sheds light on its foundational elements and historical context but also paves the way for future advancements and applications that embody the principles of intelligence in a tangible, impactful manner.

2 Core Technologies and Architectures

2.1 Sensory Integration and Multimodal Perception

The cornerstone of embodied artificial intelligence (AI) involves the adept integration of multimodal sensory data to form a cohesive understanding of dynamic environments. This section delves into the mechanisms and architectures that underpin sensory integration and multimodal perception in embodied systems. By leveraging data from various sensory modalities, such as visual, auditory, and tactile inputs, these systems aim to enhance perception capabilities, ultimately improving decision-making and interaction with the real world.

Embodied AI systems capitalize on sensor fusion techniques to effectively combine distinct sensory inputs into a unified perceptual model. Techniques like Bayesian data fusion, Kalman filters, and neural networks are widely employed for this purpose. Bayesian approaches offer probabilistic models that allow the system to update its beliefs regarding the state of the environment as new data become available, thus accommodating uncertainty and noise within sensory observations. Kalman filters, widely used in robotics for sensor fusion, provide predictions and estimates by filtering noise and inaccuracies from input data streams. More recently, the adoption of deep learning methods, such as convolutional neural networks (CNNs), is enhancing sensor fusion by learning complex feature representations from multimodal inputs [11].

However, the integration of multimodal data into a coherent perceptual framework is fraught with challenges. One significant obstacle is the alignment of asynchronous sensor data streams. Information from different modalities often arrives with varying latencies, requiring sophisticated algorithms to align and synchronize these data streams for effective processing [12]. Addressing the temporal discrepancies is crucial for achieving reliable real-time perception and is an active area of research. Techniques such as temporal convolutional networks and recurrent neural networks (RNNs) can be used to align these data streams and maintain temporal coherence in perception models.

Moreover, perception models must convert raw sensory data into actionable information. Traditional perception models use hand-crafted features, but these lack adaptability. The advent of deep learning techniques has significantly advanced unseen image recognition by allowing systems to self-optimize for feature extraction. State-of-the-art methods now employ neural models like Mask R-CNN for visual recognition tasks, where the model can segment and classify objects, even suggesting strategic paths for agents to improve recognition performance [11].

Despite these advancements, several unresolved issues remain. For instance, the handling of occlusions and contextually rich environments can still perplex embodied systems. Here, methods adopted from computational cognitive science, such as embodied cognitive frameworks, help agents interpret complex scenarios by using strategies akin to human cognitive processes, thus enhancing the contextual understanding of agents' surroundings [13].

Integration also involves real-time processing architectures, essential to the perceptual capabilities of embodied agents. Real-time processing capabilities allow agents to react instantly to environmental changes, critical for dynamic and unpredictable environments. Utilizing high-performance computing environments and optimized algorithms, embodied AI systems can achieve real-time performance. Frameworks like simultaneous localization and mapping (SLAM) have benefited immensely from parallel processing architectures and memory optimization, facilitating the development of sophisticated perception models that are both accurate and efficient [7].

Looking ahead, a synthesis of sensor fusion, perception models, and real-time processing must consider the integration of learning and adaptation. One emerging trend is the use of self-supervised learning methods that allow agents to learn directly from interactions with their environment, thus incrementally improving perception models over time without extensive external annotations [14]. Additionally, there is potential for leveraging neuromorphic models inspired by biological neural systems to develop more energy-efficient and adaptive perception architectures [15].

Moreover, as embodied AI systems become more sophisticated, the trade-off between resource constraints and perception capabilities cannot be overlooked. Aspects such as computational cost, energy efficiency, and processing speed are critical factors in the deployment of embodied AI systems in real-world applications. Addressing these concerns involves exploring hardware-specific optimizations and leveraging architectural innovations to support advanced sensory integration processes [16].

In conclusion, the sensory integration and multimodal perception capabilities fundamental to embodied AI systems are rapidly advancing through cutting-edge research and technological innovations. By combining enhanced sensor fusion techniques, sophisticated perception models, and real-time processing capabilities, embodied systems are poised to revolutionize interactions with physical environments. Future research directions include developing more adaptive and efficient models and expanding the real-world applicability of these systems. As these technologies mature, they hold the potential to advance the reach and effectiveness of embodied AI in myriad domains, from robotics and healthcare to autonomous systems and beyond.

2.2 Motor Control and Actuation Systems

Motor control and actuation systems form the cornerstone of embodied AI, translating sensory inputs into coordinated physical actions, thus bridging sensory integration and decision-making processes. These systems are essential for developing agents capable of effectively interacting with dynamic environments. This subsection examines the mathematical models, control strategies, and actuator technologies that facilitate these processes, with an emphasis on the evolving trends and challenges defining motor control within embodied AI.

At the core of motor control lie kinematic and dynamic modeling, serving as the basis for understanding and predicting movement. Kinematic models focus on the geometry of movement, employing equations such as forward and inverse kinematics to describe the configuration of robotic arms or autonomous agents without considering forces. Conversely, dynamic modeling incorporates factors such as mass, inertia, and external forces to offer a comprehensive picture of movement mechanics [3]. Understanding body schema and forward models is crucial for anticipating sensorimotor interactions and predicting future states based on current sensory data, as discussed in [3].

Adaptive control strategies play a pivotal role in uncertain and unstructured environments, allowing systems to adjust actions based on environmental feedback. These strategies, unlike static models, enhance robustness and precision in task execution. Model predictive control (MPC) and reinforcement learning strategies have gained prominence, enabling agents to anticipate changes and adjust actions dynamically. Frameworks like Distributed Adaptive Control integrate neural networks to adjust policies dynamically for embodied agents, highlighting the interplay between sensory inputs and physical actions [2]. The concept of "cheap control", which leverages an agent's embodiment for efficient behavior approximation, showcases an innovative approach, drastically reducing model complexity to achieve desired motor tasks [5].

Significant advancements in actuator technologies have enhanced the physical capabilities of embodied systems. Traditional actuators, such as electric motors and hydraulic pistons, remain foundational, yet novel technologies like artificial muscles and soft robotics, utilizing electroactive polymers or shape memory alloys, provide muscle-like capabilities with high efficiency and flexibility. These innovations have the potential to revolutionize robotic actuation, enabling smoother, more human-like movements by mimicking biological efficiencies and impacting fields like manipulative tasks in unstructured environments [17].

The integration of tactile feedback with motor control systems offers another dimension of adaptability and precision. By employing visuo-tactile perception networks, robots can enhance object manipulation skills even amidst occlusions or incomplete sensory input [18]. This cross-modal sensory integration not only heightens control accuracy but enriches the agent's interaction with its environment, enabling informed real-time decision-making despite sensory challenges.

The convergence of these technologies and control strategies is propelling the field towards developing systems that not only perform tasks effectively but also operate with efficiencies resembling biological organisms. Notable emerging trends include the integration of neuromorphic computing, which facilitates real-time processing speeds and reduced energy consumption by emulating neural architectures. Furthermore, leveraging magnetic resonance imaging technology provides robust insights into motor tasks, supporting complex decision-making processes and enhancing interaction capabilities.

Looking ahead, the primary challenge lies in crafting adaptable control systems that balance generalization and specialization across varied tasks. Developing hybrid control systems, which combine adaptive control's flexibility with rigid framework-driven approaches, represents a promising research direction. Additionally, incorporating active inference models to minimize prediction errors and boost embodied intelligence presents a significant opportunity for future innovation [19].

In conclusion, the field of motor control and actuation in embodied AI is rapidly evolving, offering tremendous potential for innovation and application. By integrating advanced models of motor coordination with cutting-edge actuator technologies, the future of embodied AI promises increasingly sophisticated systems capable of harmonious operation within their environments. Continued interdisciplinary collaboration across robotics, neuroscience, and machine learning will drive the development of embodied systems that are more adaptable, efficient, and versatile, thus elegantly addressing challenges with the same finesse characteristic of natural organisms.

2.3 Cognitive Architectures and Decision-Making

In the realm of Embodied Artificial Intelligence (EAI), cognitive architectures serve as the blueprint for complex decision-making processes, mirroring human-like reasoning and adaptability. This subsection focuses on dissecting these frameworks with an emphasis on symbolic and sub-symbolic integration, emotion and social cognition integration, and hierarchical control models.

At the heart of decision-making in EAI lies the integration of symbolic and sub-symbolic processes. Symbolic systems are traditionally associated with rule-based logic and structured problem-solving, which provides clear, interpretable decisions. Sub-symbolic systems, often realized through artificial neural networks, excel in capturing complex patterns from vast datasets but are typically opaque in their decision-making rationale. The integration of these paradigms seeks to harness the strengths of both systems for holistic decision-making processes. For instance, the TAME framework, as discussed in [20], formalizes a non-binary approach to strongly embodied agency, demonstrating how continuous, empirically-based subsymbolic processes can be embedded into structured symbolic paradigms. This dual approach enables EAI systems to bridge the gap between explicit logical reasoning and nuanced pattern recognition, a synergy that has been leveraged to enhance task performance in dynamic environments.

The integration of emotional and social cognition into decision-making architectures is pivotal for achieving nuanced human-robot interactions. Embodied AI systems must interpret emotional cues and social contexts to interact effectively in human-centric environments. The inclusion of emotional cognition can significantly influence decision-making, akin to human biases in decision processes. In [21], the integration of neuromorphic computing models that simulate brain-like processing is proposed, potentially allowing robots to assess and respond to social cues with a refined understanding akin to human emotional interpretation. The challenge remains to model these complex emotional states accurately and to ensure these AI systems respect ethical standards in their social interactions. The purpose is to enhance collaboration with humans, striking a balance between empathetic understanding and maintaining functional autonomy in decision-making scenarios.

Hierarchical control models in EAI provide a structured mechanism for decision-making, where strategic planning operates across several tiers of abstraction. These models decompose complex actions into manageable subtasks, which allow for high-level decision-making that informs lower-level actuator commands. An illustrative example is provided in [22], where high-level task directives are converted into focused, low-level movements through a multi-layered control system. Hierarchical approaches facilitate adaptability and robustness in AI decision-making, offering a strategic advantage over flat-controlled systems which may struggle with the intricacies of real-time environmental feedback.

The comparative efficacy of these cognitive architectural paradigms illustrates diverse benefits and limitations. Symbolic systems, while easy to validate and interpret, lack the adaptability and pattern recognition capabilities of sub-symbolic systems. Conversely, sub-symbolic networks require substantial training data and can encounter difficulties in providing interpretable outputs [23]. Emotion and social cognition integration offer profound advancements in human-machine interaction but introduce complexity in ensuring reliable, context-appropriate interactions. Hierarchical models excel in providing structured and layered decision-making but may face challenges in real-time adjustments and computational demands.

Emerging products in this domain, such as [6], highlight technological trends towards the co-evolution and optimization of both cognitive architectures and physical embodiments. Moreover, with the proliferation of neuromorphic processing platforms [24], there is a growing exploration of brain-inspired computing in decision-making processes, leveraging the natural synergies of perception, cognition, and physical interaction in embodied systems.

As we venture into more complex and human-like decision-making architectures, several challenges demand attention. These include the development of ethical frameworks guiding decision-making processes in socially sensitive environments, ensuring transparency and accountability in AI decisions, and paving pathways for real-time learning and adaptability without explicit human intervention. As noted in [25], exploiting morphological computation for decision-making can significantly reduce controller complexity by offloading cognitive computations to physical interactions, but necessitates innovative control strategies that reliably map sensory feedback to actuator commands without extensive computational overheads.

In conclusion, the exploration of cognitive architectures in decision-making remains a rich area for innovation, with substantial scope for improving interaction quality between embodied AI systems and their environments. Future research directions could benefit from integrated, cross-disciplinary approaches drawing inspiration from neuroscience, cognitive science, and robotics, focusing on developing systems capable of seamless adaptation and contextual understanding in unstructured, real-world environments. Such advancements will likely drive the next wave of AI systems, thereby transforming embodied intelligence into a multifaceted and omnipresent reality.

2.4 Embodiment in Multimodal Interaction

Embodiment in multimodal interaction is a crucial pillar of embodied artificial intelligence (AI), utilizing integrated sensory and actuation frameworks to enable more human-like engagements. This subsection delves into these interactions, discussing the convergence of voice, gesture, and other modalities that enhance communication and collaboration within embodied AI systems, thereby acting as a bridge between the cognitive architectures previously discussed and the simulation and development platforms that follow.

Multimodal dialogue systems are a central component within this realm, recognized for their sophisticated integration of numerous inputs such as auditory, visual, and tactile information. These systems strive for seamless interaction models similar to human dialogues, which naturally incorporate varied sensory cues. By synchronizing modalities—like speech and gesture—a robot can respond with contextually appropriate actions, enhancing its interaction capabilities. The DAC-h3 cognitive architecture exemplifies this intersection, where multimodal inputs are processed through biologically grounded frameworks to enable adaptive, context-sensitive actions [26]. This integrated approach highlights the need for multimodality to transcend isolated command-following, moving towards more fluid conversational exchanges poised for real-world applicability.

Affective computing, another pivotal element, focuses on systems' capacity to recognize and interpret emotional cues, promoting empathetic and adaptive interactions. Robots equipped to perceive emotions can customize their responses to human affective states, as demonstrated by systems like Ryan—a robot designed for social interaction with older adults that engages users by recognizing emotions through multimodal inputs [27]. This computational empathy, achieved via sentiment analysis and emotional interpretation of vocal and facial cues, permits more personalized and meaningful interactions, reflecting the social cognition integration discussed earlier.

Incorporating continuous context learning is integral to well-rounded multimodal systems, enabling AI to adapt behavior dynamically based on evolving situational contexts. Concepts like Theory of Mind—where AI infers human goals and intentions based on observed actions—play a vital role in this understanding. The Leolani project exemplifies this dynamic adaptation, where robots update their knowledge bases through social interactions, internalizing new information to resolve uncertainties and contradictions, marking a crucial step toward nuanced multimodal engagement [28].

Despite significant progress, these systems encounter challenges such as sensory data ambiguity and environmental noise. Neuro-symbolic approaches offer promising solutions, combining deep learning capabilities with reasoning and structured knowledge representation to address these complexities [29]. Such hybrid models can better navigate sensory data intricacies, providing enhanced robustness and interpretability, aligning with the cognitive architectural models previously discussed.

A comparative analysis reveals distinct trade-offs within current approaches. Systems heavily reliant on neural architectures excel in learning and adapting from extensive datasets through direct sensory input but may struggle with interpretability and abstract reasoning. Conversely, systems based on symbolic processing provide strong deductive reasoning but falter in handling unstructured, noisy inputs. Embedding hierarchical processing, as illustrated by Neural Modular Control strategies, enables layered input processing to encompass both immediate sensory perceptions and higher-order reasoning [30].

Emerging trends indicate a shift toward tightly integrated cognitive architectures that blend real-time perception with higher-order, symbolic reasoning. The evolving field of human-aware AI underscores the necessity for sophisticated multimodal frameworks, allowing systems not only to react but also predict human actions, aligning closely with user expectations and demands.

Looking forward, the development of more immersive simulation environments is anticipated to be pivotal in advancing these capabilities. These high-fidelity platforms will facilitate training and testing of embodied AI systems' multimodal interaction capabilities in complex social environments, bridging the sim-to-real gap and fostering adaptable AI systems ready for seamless real-world deployment [31].

In conclusion, while multimodal techniques in embodied AI continue to promise enriched human-computer interaction, further research is essential to refine these systems' contextual understanding and empathetic capabilities. The ongoing synthesis of deep learning and neuro-symbolic techniques, supported by advanced simulation platforms, is poised to drive future advancements, advancing AI that truly reflects the nuances of human communication and collaboration.

2.5 Simulation and Development Platforms

Simulation and development platforms are integral to the advancement of embodied AI, playing a critical role in designing, testing, and iterating prototypes before deploying them in real-world environments. These platforms offer virtual environments where embodied AI systems can be developed and tested in a controlled setting, which aids in overcoming challenges such as ensuring safety, minimizing cost, and handling the complexities of real-world interactions. This subsection explores various platforms and methodologies, comparing their strengths, limitations, and emerging trends, while also highlighting their significance in education and training contexts.

The primary function of simulation platforms in embodied AI is to provide high-fidelity environments that replicate real-world scenarios with sufficient accuracy to train systems robustly. High-fidelity simulators, such as EmbodiedScan, are examples of platforms tailored for developing embodied AI by offering detailed 3D scene understanding and dynamic interaction capabilities [32]. These platforms allow for comprehensive sensory and motor modeling, enabling the exploration of complex task executions that are critical for developing autonomous agents capable of nuanced interactions.

An essential aspect of these platforms is their ability to facilitate the sim-to-real transfer, a crucial process in bridging the gap between virtual training and real-world deployment. The reality gap—discrepancies between simulation models and real-world physics—poses a significant challenge, as systems can perform well in virtual environments but fail when exposed to the unpredictability of the real world [33]. Domain randomization and physics-informed models, such as those discussed in MotionChain, are emerging techniques that address this issue by promoting generalization and robustness, allowing agents to adapt to complex and varying environments [34].

Beyond providing a testing ground for systems, simulation platforms are invaluable for iterative development and collaborative innovation. Open-source collaborative development tools like those used in DialFRED enable researchers to share and build upon each other's work, accelerating progress in embodied AI through shared datasets and standardized protocols [35]. Such platforms not only promote innovation through collective effort but also ensure consistency and replicability in experimental designs, making it easier for the community to compare methodologies and outcomes.

While the advantages of simulation platforms are significant, they are not without limitations. High-fidelity simulators often require substantial computational power, which can be a barrier to widespread accessibility and scalability. Additionally, not all aspects of human-computer interactions—such as emotional cues or social dynamics—can be seamlessly integrated into simulations, limiting their ability to fully replicate every nuance of real-world scenarios [36]. Despite these challenges, ongoing advancements in multimodal data integration and computational efficiency continue to improve the fidelity and accessibility of such platforms.

In the educational realm, these platforms offer powerful tools for teaching and training. By allowing students and practitioners to experiment with embodied AI systems in a risk-free environment, platforms like EmbodiedScan provide experiential learning opportunities that are invaluable for developing practical skills and understanding complex AI concepts [32]. They also serve as a springboard for innovation, encouraging learners to explore novel approaches and solutions without the constraints typically encountered in physical settings.

Looking forward, the trend in simulation and development platforms is moving towards greater interoperability and multimodality, influenced by advances in areas such as large multimodal agents (LMAs) and multi-modal large language models (MLLMs). These technologies leverage rich sensory data to create more holistic and context-aware simulations that better reflect real-world complexities [37]. Developments in virtual and augmented reality (VR/AR) also promise to enhance the immersive quality of simulations, offering more realistic and intuitive interfaces for interaction and control.

In conclusion, simulation and development platforms are pivotal to the progress of embodied AI, offering indispensable tools for testing, iteration, and education. As technologies advance, these platforms are expected to evolve towards more sophisticated, multimodal environments that closely mimic real-world complexities, thereby improving the effectiveness of AI training and deployment. By facilitating a deeper understanding of systems and fostering collaborative innovation, these platforms pave the way for the widespread adoption and advancement of embodied AI technologies. The pursuit of bridging the reality gap, combined with enhanced collaborative frameworks, will shape the future trajectory of this exciting field.

3 Learning Mechanisms in Embodied Systems

3.1 Paradigms of Reinforcement and Imitation Learning

Within the realm of embodied artificial intelligence, reinforcement learning (RL) and imitation learning (IL) emerge as pivotal paradigms that enable agents to acquire complex behaviors through interaction and observation, respectively. These paradigms not only facilitate robust decision-making in dynamic environments but also allow agents to emulate human-like behaviors effectively. The integration of RL and IL within embodied systems has led to groundbreaking advancements in autonomous robotics, adaptive control, and human-robot interaction, embodying both challenges and opportunities for future exploration.

Reinforcement Learning (RL): At its core, RL is a goal-oriented learning paradigm where an agent learns optimal policies through trial and error by interacting with its environment to maximize cumulative rewards. In the context of embodied systems, RL has been employed to address complex control tasks, such as locomotion, manipulation, and navigation. For instance, the Deep Evolutionary Reinforcement Learning (DERL) framework has demonstrated the potential of evolving diverse agent morphologies to learn locomotion and manipulation tasks autonomously, underscoring the importance of RL in fostering morphological intelligence within complex environments [38].

The underlying mechanism of RL operates on the principles of state, action, and reward, formalized through Markov decision processes (MDPs). The agent, being in a particular state, selects an action based on a policy, receiving a reward and transitioning to a new state. The objective is to find a policy that maximizes expected reward over time, encapsulated mathematically by optimizing the value function using algorithms like Q-learning, policy gradient methods, and actor-critic models. These approaches vary in their capacity to handle issues like exploration, sample efficiency, and function approximation, each with specific trade-offs [39].

Despite its efficacy, RL faces substantial challenges in real-world embodied systems. The high dimensionality and stochastic nature of real-world environments often result in sparse or delayed rewards, rendering traditional RL methods inefficient. Moreover, RL can be sample-inefficient, requiring substantial interaction data, which is impractical for real-world deployment. Techniques such as reward shaping, hierarchical RL, and intrinsic motivation are being explored to mitigate these bottlenecks, enhancing the scalability and applicability of RL in embodied contexts [40].

Imitation Learning (IL): In contrast to RL, imitation learning (IL) involves learning from demonstrations, allowing agents to acquire complex behaviors by observing expert actions. This paradigm bypasses the trial-and-error aspect of RL, offering a more sample-efficient approach to acquiring policies directly from observed trajectories. Behavioral cloning and inverse reinforcement learning (IRL) are central methods within IL, each with distinct methodologies. Behavioral cloning involves supervised learning techniques to map states directly to actions, whereas IRL attempts to infer the underlying reward structure from demonstrations to derive optimal policies [41].

IL's efficacy in embodied systems shines in scenarios where expert demonstration data are readily available or when learning a new behavior rapidly is crucial. However, a core challenge in IL lies in generalizing from limited demonstrations, particularly in novel environments. Recent advancements have focused on using IL in conjunction with reinforcement learning to capitalize on their respective strengths—a hybrid approach wherein IL is initially used to bootstrap policies, which are subsequently refined via RL as the agent interacts with its environment [2].

Integration of Human Feedback: A transformative avenue in RL and IL is the integration of human feedback, which leverages human evaluators to guide the learning process. Reinforcement Learning from Human Feedback (RLHF) stands out as an approach combining human oversight with autonomous learning to refine agent behavior, aligning it more closely with human expectations. By incorporating qualitative assessments from humans, agents can calibrate their decision-making processes more effectively, enhancing both safety and ethical compliance in human-centric applications such as assistive robotics and social interaction [42].

Moreover, the co-evolution of agents with interactive human-centric environments, drawing from enaction-based paradigms, emphasizes the reciprocal influence of environment and agent, providing a more holistic development framework for embodied systems [13]. Future explorations in integrating enactive intelligence into RL and IL could enable more seamless adaptations to human social norms and nuances.

Trends and Future Directions: The ongoing evolution of RL and IL within embodied systems indicates a trajectory towards more integrative and multimodal learning approaches. The potential for transferring simulated learnings to real-world applications via advancements in sim-to-real transfer methods remains a key research pursuit. Simultaneously, the exploration of lifelong and continual learning frameworks as an adjunct to RL and IL promises to address the limitations of static policy models, fostering adaptability to evolving contexts and missions [9].

In conclusion, while RL and IL present distinct pathways toward realizing efficient decision-making and human-like behavior in embodied systems, their integration and co-evolution with human-centric feedback and environmental dynamics are poised to redefine the capabilities of artificial agents. Bridging these paradigms with emerging technologies like neuromorphic computing and bio-inspired designs could unlock unprecedented potentials for embodied intelligence, paving the way for transformative societal and technological impacts.

3.2 Transfer Learning and Cross-Domain Adaptation Techniques

Transfer learning and cross-domain adaptation are pivotal mechanisms through which embodied systems gain the flexibility to operate across varied environments and tasks, thereby enhancing both their scalability and robustness. This subsection delves into the methodologies by which embodied artificial intelligence systems employ these techniques, the comparative advantages they offer, their limitations, and emerging trends in the field.

Transfer learning refers to the process of leveraging prior knowledge or trained models from one domain or task to improve learning efficiency and performance in a different, often related, domain or task. In the context of embodied systems, which encompass robots and agents capable of interacting with the physical world, the ability to transfer knowledge is critical for efficient adaptation to new environments without requiring extensive retraining. The fundamental challenge lies in effectively mapping knowledge from the source domain (where data or experiences are rich) to the target domain (where data may be sparse or differ significantly).

1. **Knowledge Transfer Modality:** Frameworks facilitating knowledge transfer in embodied systems often focus on transferring components such as policies, value functions, or dynamics models across domains. These include model-based transfer learning, where a dynamics model learned in one environment is adapted for use in another. Techniques explored in [43] demonstrate the adaptability of high-level policies in new domains through retraining on reduced datasets.

2. **Domain Adaptation Strategies:** Domain adaptation focuses on aligning distinct domain representations to facilitate learning transfer. Sim-to-real transfer, a key area in robotics, explores strategies for deploying policies learned in simulation environments to real-world applications, addressing discrepancies known as the 'reality gap'. Techniques like domain randomization, where environmental variables are randomized to improve policy robustness, are widely used. Effective sim-to-real adaptations are discussed in papers like [44], showing the efficacy of sensorimotor learning in simulation before real-world deployment.

3. **Meta-Learning Approaches:** Meta-learning, or "learning to learn", enhances dynamic transfer learning by training models to rapidly adapt to new tasks using prior knowledge. This approach treats the learning algorithm itself as an optimization problem, enabling quick adaptation to new scenarios by updating only a few parameters. Methods that emphasize such rapid adaptation have been advanced in works like [45], highlighting the importance of embodiment in learning symbolic representations quickly.

The main strength of transfer learning in embodied systems is its ability to reduce the data and computational requirements for training models in new environments. Instead of starting from scratch, systems can build upon already acquired skills, leading to faster convergence in new task settings. This aspect is crucial for applications where computational resources or opportunities for data collection are limited. However, limitations exist primarily in areas where source and target domains have significant variability in dynamics or sensory inputs, making direct transfer inefficient or leading to performance degradation.

An emerging trend is the integration of transfer learning with reinforcement learning frameworks to enhance robotic autonomy. Autonomous agents benefit from transfer learning by using previously gathered experiences to inform decision-making processes in novel but structurally similar contexts, as seen in [46]. Moreover, the convergence of transfer learning with large language models (LLMs) exemplifies promising interdisciplinary innovations, where pre-trained models enhance embodied decision-making processes by retrieving richly encoded representations of visual and language data, as discussed in [47].

The practical implications of effective transfer learning and domain adaptation strategies are profound, particularly for industry applications such as autonomous vehicles, healthcare robots, and domestic assistance systems. These systems require robust adaptation to diverse real-world conditions, a capability significantly enhanced by transfer learning techniques.

Looking forward, future directions of this field are poised to focus on enhancing the granularity of transfer techniques to support broader and more intricate task domains. This includes fine-grained transfer that not only maps overarching policies but also adapts intricate skill patterns and context-aware behaviors. Research innovations may explore hybridized transfer learning frameworks that integrate aspects of human cognitive flexibility, allowing for meta-cognitive adaptations leveraging diverse multimodal sensory inputs, as outlined in [3].

In conclusion, transfer learning and cross-domain adaptation are not merely supportive tools but foundational components of embodied AI, holding immense potential to propel the field towards more adaptable and intelligent systems capable of operating seamlessly across a vast array of environments and tasks. This transformative potential hinges on continuous advancements in aligning theoretical frameworks with practical implementations, enabling cognitive architectures in robots and agents to realize their full capability in versatile applications.

3.3 Continual and Adaptive Learning Frameworks

Continual and adaptive learning frameworks are crucial for embodied AI systems to thrive in dynamic and evolving environments. These frameworks enable embodied agents to accumulate knowledge incrementally and adapt to new contexts without the degradation of previously acquired skills, known as catastrophic forgetting. This subsection explores various strategies and models designed to achieve continual learning in embodied AI, analyzing their effectiveness, limitations, and potential for future development.

One of the foundational concepts in continual learning is lifelong learning, where agents progressively build on their knowledge base throughout their operational lifespan. Traditional machine learning models often struggle with catastrophic forgetting, where new learning leads to the erosion of previously learned information. Various techniques have been developed to address this issue, such as regularization-based approaches, which introduce penalty terms during the update of model parameters to preserve previously learned knowledge. For instance, Elastic Weight Consolidation (EWC) and its variants use Fisher Information Matrix to selectively penalize changes to important model parameters when learning new tasks, maintaining a balance between adaptability and retention of past expertise.

Online and incremental learning algorithms play a critical role in enabling agents to learn from continuous data streams, updating their models in real time. These algorithms can be particularly beneficial in embodied systems, where sensory information is constantly streaming, and the environment is perpetually changing. Incremental learning frameworks are designed to process each data point or mini-batch sequentially, reframing learning as an ongoing process rather than periodic updates based on fixed datasets. These methodologies leverage approximation techniques and scalable architectures to handle vast data efficiently, allowing agents to remain responsive and adaptable [24].

Adaptive behavior formulation and strategic planning are also pivotal in continual learning frameworks. Embodied agents must not only learn new tasks but also develop adaptive strategies to manage varying environmental conditions and task demands. This is exemplified by approaches that integrate predictive models to anticipate future states and plan actions accordingly [48]. In such models, predictive state representations can inform the decision-making process, aligning actions with anticipated changes in environmental context. Moreover, learning paradigms such as Reinforcement Learning (RL) and Model Predictive Control (MPC) are often integrated into these frameworks to provide a structured approach to strategy formulation. By integrating a feedback loop from the environment and continuously updating policy learning, these systems can maintain resilience and adaptability in even the most unpredictable scenarios [49].

Another crucial aspect of continual learning frameworks is the ability to facilitate transfer learning between tasks and environments. Transfer learning techniques aim to generalize knowledge acquired from one domain to others, thereby enhancing the versatility and efficiency of learning processes. Domain adaptation and task invariance methodologies have shown promise in maintaining performance across differing environments by accounting for domain-specific characteristics [50]. For example, the transfer of locomotion skills from simulation to real-world settings often involves compensating for discrepancies in dynamics through sim-to-real adaptation techniques [21].

However, several challenges persist in the implementation of continual and adaptive learning in embodied systems. One of the primary challenges is ensuring the stability and scalability of learning algorithms to prevent divergence and ensure robustness over extended deployments. Scalability remains a significant barrier, particularly as embodied systems operate across larger and more complex state spaces, each requiring tailored strategies for effective learning [51]. Additionally, maintaining a consistent level of adaptiveness without succumbing to catastrophic forgetting remains a pivotal area of research, with methods such as replay memories and dual-learning networks being explored to mitigate these challenges.

Emerging trends in continual learning frameworks focus on more holistic and integrated approaches, combining insights from neuroscience, cognitive science, and machine learning to inform embodied agent design. Concepts such as hierarchical reinforcement learning, which employs a tiered structure to encompass both high-level strategic decision-making and low-level perceptual control, are gaining traction. These models provide a structured pathway for managing the complexity of continual learning by compartmentalizing tasks at different hierarchical levels, thus facilitating adaptability and cognitive scalability [22].

In conclusion, continual and adaptive learning frameworks represent a dynamic and expanding field within embodied AI, holding immense potential for enabling agents to function in constantly evolving environments. Although significant progress has been made, challenges such as catastrophic forgetting, scalability, and effective transfer learning require ongoing research and innovation. By synthesizing advancements in various learning paradigms and drawing upon interdisciplinary insights, future research endeavors can enhance the reliability, adaptability, and cognition of embodied systems. This ongoing evolution promises not only to advance the capabilities of artificial agents but also to deepen our understanding of embodied cognition as a whole.

3.4 Multimodal Learning and Integration in Embodied Systems

The integration of multimodal sensory inputs within embodied AI systems is a pivotal frontier that enhances the cognitive capabilities of artificial agents. Methodologies and innovations enabling these systems to efficiently process heterogeneous data streams—spanning visual, auditory, tactile, and proprioceptive inputs—formulate a cohesive understanding of their environment. This improved capability engenders enhanced interaction with the environment and decision-making faculties, both critical for achieving nuanced, human-like cognitive processes in artificial agents.

As embodied systems increasingly operate in dynamic and unstructured environments, they must process multiple sensory inputs simultaneously. Sensor fusion techniques provide a critical framework for synthesizing data from diverse modalities into a unified perceptual model. For instance, integrating visual and proprioceptive inputs can significantly enhance spatial awareness and navigation capabilities. Additionally, embodied question answering leverages hierarchical policies to effectively use language inputs, demonstrating a method for aligning complex sensory data with action tasks [30].

The technical strategies supporting sensor fusion include probabilistic models, tensor-based fusion architectures, and neural-symbolic integration. Tensor-based approaches facilitate high-throughput multidimensional data processing, capturing complex relationships across sensory modalities with precision. Probabilistic models offer a framework for handling the inherent uncertainty of sensory data, thus improving robustness against noise and errors. However, achieving an optimal balance between computational efficiency and sensory input precision remains a crucial challenge [52].

A significant feature of multimodal learning is its facilitation of cross-modal knowledge transfer, enabling learning in one sensory domain to support another. For example, knowledge gained from visual data can enhance auditory processing, providing an enriched foundation for reasoning and adaptation across various contexts. This transfer is essential for maintaining functionality when one sensory channel is impaired or unavailable. Techniques such as adversarial training within multimodal neural architectures have advanced this cross-modal sharing, boosting both resilience and adaptability in sensory-deprived scenarios [29].

Cross-modal learning systems often utilize unsupervised and semi-supervised learning paradigms, facilitating complex environmental models without extensive labeled data. These approaches are especially valuable as embodied systems transition from controlled to real-world environments where labeled datasets may be sparse. Implementing such paradigms mitigates data bottlenecks and enables smoother adaptation to diverse tasks and environments [53].

The application of multimodal sensory inputs significantly enhances task performance across various domains. Systems employing multimodal data show improvements in tasks such as navigation, manipulation, and interaction compared to unimodal systems. In social robotics, for instance, combining visual, auditory, and tactile data refines interpretative capabilities, allowing robots to recognize and empathetically respond to human emotional states—a crucial step toward socially intelligent machines [54].

Despite these advancements, several challenges persist in multimodal integration, notably designing efficient system architectures that can process multimodal data in real-time. Achieving scalable solutions applicable across various hardware configurations and environments remains daunting [55]. Additionally, as systems mimic human cognition, ethical considerations such as user privacy and security within complex datasets are paramount.

Looking ahead, the evolution of multimodal learning in embodied systems will likely be guided by advances in neurosymbolic AI, merging symbolic reasoning with neural network learning paradigms. This evolution aligns with the imperative for these systems to not only interpret but also communicate and justify their actions, marking a shift toward symbiotic human-machine interactions.

In conclusion, multimodal learning marks a significant advancement toward more adaptive and intelligent embodied systems. Overcoming the limitations of traditional unimodal approaches promises enhanced interaction competencies, facilitating more effective deployment in complex, real-world scenarios. Future research must address the computational and ethical challenges of multimodal systems, driving breakthroughs that unlock the full potential of embodied artificial intelligence.

3.5 Ethical and Societal Implications of Learning Mechanisms

In the burgeoning field of Embodied Artificial Intelligence (AI), the learning mechanisms that enable these systems to perform autonomously bring along profound ethical and societal implications. As these systems evolve, embedding them with learning capabilities—ranging from reinforcement and imitation learning to continual and multimodal learning—demands careful consideration of issues like transparency, accountability, societal acceptance, and regulatory frameworks. This subsection delves into the ethical and societal aspects of learning mechanisms in embodied AI, examining the challenges these systems pose and exploring potential future directions for addressing them.

The deployment of embodied AI systems necessitates a rigorous evaluation of ethical challenges, particularly around privacy and bias. These systems often collect vast amounts of sensory data from their environments, raising concerns about user privacy. Ensuring that sensitive data are not misused requires embedding privacy protection measures within the AI's core architecture. For instance, strategies involving data anonymization, encryption, and clear user consent protocols must be established to safeguard personal information [56].

Bias in learning mechanisms is another pertinent ethical issue. The models and algorithms guiding these systems must be analyzed for inherent biases that may arise from skewed training datasets or biased reinforcement strategies. It is critical to implement fairness-centered design approaches that can detect and mitigate these biases to prevent the perpetuation of societal stereotypes and unfair treatment [57]. Furthermore, transparent decision-making processes should be prioritized to enhance accountability, allowing end-users and developers to audit and understand the AI's actions and learning trajectories.

In the context of societal acceptance, it is crucial to navigate the intricate dynamics of human-AI interaction. For embodied systems to be integrated effectively within societal frameworks, they must align with social norms and cultural contexts. Perception plays a pivotal role in acceptance, as systems perceived as trustworthy and useful are more likely to be embraced. Enhancing embodied AI's emotional intelligence can aid in this, aligning the AI's responses with societal expectations, as evidenced by [56], which emphasizes the empathy module's role in interpreting and responding to human emotions.

Adaptive human-embodied system collaboration necessitates learning mechanisms that do not merely imitate human actions but also respect and adapt to human intentions and social cues. As illustrated in [58], empathy and social awareness are of paramount importance in shaping nuanced and culturally aware interactions. Moreover, to address privacy concerns and improve trust, incorporating explainability and user-centric design principles in embodied AI's learning frameworks becomes imperative [57].

Safety is an overarching concern when deploying AI systems that learn and operate autonomously. Establishing safety protocols and regulatory guidelines is essential to mitigate risks associated with unpredictability in autonomous learning. Studies such as [59] have demonstrated that integrating robust safety measures during the design phase of embodied systems can significantly reduce erroneous actions that might endanger users.

Regulating learning paradigms involves defining verifiable performance metrics and establishing benchmarks to ensure that embodied systems' operations align with safety and ethical standards. Regulatory bodies should consider the specific risks presented by the complex and adaptive nature of embodied AI, implementing rigorous testing and certification processes. Structured frameworks like those discussed in [60] provide a foundation for evaluating social interactions within AI systems, fostering a regulated approach to deployment.

In synthesizing these considerations, it becomes evident that the future of learning mechanisms in embodied AI hinges upon our ability to balance technical advancements with ethical and societal accountability. One promising direction is the development of modular frameworks that can integrate ethical guidelines directly into the learning processes, making ethics an intrinsic part of the AI's decision-making process. Additionally, fostering interdisciplinary collaborations among AI researchers, ethicists, policymakers, and societal stakeholders is essential for developing holistic solutions to these multifaceted challenges [61].

Looking forward, efforts should focus on enhancing the transparency of learning mechanisms and achieving a symbiotic relationship between humans and embodied AI systems. As we strive towards this vision, it is crucial to keep refining the frameworks governing autonomous learning to ensure these systems not only perform optimally but also resonate with human values and ethical considerations. By doing so, we can pave the way for embodied AI systems that are not only technologically advanced but also ethically sound and socially responsible.

4 Embodied Interaction and Human-Robot Collaboration

4.1 Fundamentals of Human-Agent Interaction

The interaction between humans and embodied AI agents is a focal point in advancing the usability and acceptance of robotic systems in diverse environments. As such, understanding the fundamentals of human-agent interaction is crucial for effective collaboration and seamless operation in shared spaces. This subsection explores the frameworks, methodologies, and technologies that have been developed to facilitate human-agent interaction, emphasizing aspects such as turn-taking, contextual awareness, and adaptive learning.

Human-agent interaction is inherently complex due to the dynamic nature of human behavior and the evolving expectations from AI systems. One foundational element is the development of structured interaction protocols, which provide a scaffolding that governs the communication patterns between agents and humans. These protocols encompass turn-taking methods, signaling mechanisms, and error correction strategies that ensure mutual understanding and efficient task execution [62]. In considering interaction patterns, it is vital to incorporate adaptive elements that account for variations in human intentions and contextual cues, thereby promoting a more intuitive interaction experience.

Traditional interaction models have frequently utilized rule-based systems which rely on predefined scripts and dialogue trees. However, these models have limitations in adapting to spontaneous human behaviors and novel situations. Recent advancements have focused on more flexible and adaptive interaction frameworks that employ machine learning techniques to infer human intentions and adapt responses accordingly [13]. These systems leverage historical data and context-aware algorithms to dynamically adjust their strategies, ensuring that interactions remain fluid and relevant to the task at hand.

A key challenge in enabling effective human-agent interaction lies in understanding the subtleties of human communication, including non-verbal cues such as gestures, facial expressions, and body language. These cues provide an additional layer of information that can significantly enhance the interaction process if effectively interpreted by embodied agents. For instance, the integration of computer vision systems with emotion recognition technologies allows robots to assess the emotional states of their human partners, adjusting their behavior to maintain harmonious interactions [63].

The role of multimodal communication frameworks in facilitating richer and more immersive human-agent interactions cannot be overstated. By integrating various sensory inputs, such as auditory, visual, and tactile signals, embodied agents can construct a more comprehensive understanding of their environment and the humans they interact with. This integration enables the development of dialogue systems that are capable of responding to both verbal commands and non-verbal cues, leading to more natural interactions [64].

Despite the significant progress in improving interaction frameworks, there remain notable challenges and trade-offs in balancing system complexity, interpretability, and real-time performance. Advanced interaction systems often require substantial computational resources, posing scalability issues especially in resource-constrained settings. Additionally, the complexity of adaptive models may obstruct the interpretability of agent actions, complicating troubleshooting and system trustworthiness [5].

Emerging trends in literature suggest a concerted effort towards incorporating the principles of embodied cognition, which emphasize the role of the physical form and sensorimotor processes in shaping cognitive functions. Recent works propose integrating real-world sensorimotor data into cognitive architectures, thereby enhancing the agents' ability to infer and reason about human actions and intentions [3].

Looking forward, the future of human-agent interaction will likely see greater focus on the integration of enactive AI principles, where interaction is seen as a form of cooperation that continuously evolves through feedback from the environment and participating agents. This aligns with the notion of participative AI, where the AI system actively learns from its interactions, adapting its behaviors and guidelines in real-time [13].

In conclusion, the field of human-agent interaction is advancing towards more intuitive, adaptive, and context-aware systems aided by multimodal sensory integration and foundational interaction protocols. Future research must focus on addressing the challenges of scalability, interpretability, and robustness of these systems to facilitate their widespread adoption and efficacy in real-world applications. Collaborative efforts in bridging gaps across AI subdomains and improving the interdisciplinary nature of research will be crucial for fostering innovations that enhance human-agent interaction dynamics.

4.2 Emotional and Social Intelligence in Human-Robot Interaction

The integration of emotional and social intelligence into human-robot interaction systems represents a significant advancement in making robotic interactions more intuitive and relatable. Building on the principles of human-agent interaction, these systems enable robots to move beyond functional engagements toward more empathetic, context-aware interactions, serving as an extension of the adaptive and multimodal communication frameworks discussed earlier. This subsection delves into the multifaceted capabilities of robots in interpreting and responding to human emotions and social cues, exploring various approaches, their implications, and future directions.

Central to these systems is the capability to recognize and interpret human emotions accurately. Robots equipped with advanced sensor arrays and machine learning algorithms can detect cues from facial expressions, body language, and vocal tones to infer emotional states. Techniques such as deep affordance-grounded sensorimotor learning allow robots to understand object "affordances" or typical interaction dynamics, which can be extended to recognize subtle emotional cues [46]. These robots utilize neural networks that mimic human perception processes, enabling them to predict and adapt to human emotional responses accurately, thus aligning with the adaptive learning models highlighted previously.

Emotion recognition systems often employ multimodal approaches to synthesize visual, auditory, and sometimes haptic feedback, echoing the multimodal communication frameworks outlined in the subsequent section. Multimodal fusion facilitates robust emotion interpretation by combining data streams, thereby reducing the ambiguity present in mono-modal systems [65]. This integration supports more sophisticated interaction paradigms, fostering empathetic and adaptive behaviors that humans naturally respond to, and enhancing the interaction quality as emphasized in both preceding and following frameworks.

Understanding social cues—such as proximity, gaze, and gestures—is essential for robots to navigate social contexts effectively. Social cue processing relies on both the structural design of AI systems and their cognitive architectures. An emerging approach involves symbolic and sub-symbolic integration within these architectures, leveraging both linguistic and non-linguistic inputs to infer social dynamics [4]. By associating these social indicators with potential actions, robots can better conform to social norms and expectations, enhancing their acceptability and integration into human environments, a theme that resonates with the creation of seamless human-agent communication pathways.

Designing robots with empathy remains a complex task, requiring sophisticated models capable of contextually relevant responses. Enhanced robot speech recognition systems that combine binaural sound source localization and advanced natural language processing enable robots to engage effectively in conversations, even in noisy environments, by adjusting to explicit and implicit emotional content [66]. Additionally, frameworks like the joint model of language and perception play a crucial role in grounding linguistic expressions in perceptual contexts, offering robots nuanced understanding and response capabilities [67].

The broader impact of socially intelligent robots extends beyond surface-level interaction improvements. Empathic robots hold the potential for profound applications in healthcare, education, and eldercare, where emotional intelligence is vital. These robots can offer companionship or assist with therapy by understanding and responding to the emotional needs of users, improving psychological and emotional well-being. However, deploying such systems presents challenges related to privacy, security, and societal ethics. Robots with access to sensitive emotional data need robust privacy protections and transparent operations to ensure trust and acceptance by users, reflecting an overlap with ethical considerations highlighted in interaction dynamics.

As research progresses, there is a growing recognition of the complexities involved in seamlessly embedding emotional and social competencies into robotic systems. Future research directions involve not only refining the technical capabilities of robots but also addressing broader questions about the nature of empathy in machines, ethical interaction designs, and the societal impact of widespread adoption of socially intelligent systems. The evolving landscape of emotion-driven AI opens avenues for interdisciplinary collaborations, integrating insights from psychology, cognitive science, and artificial intelligence to build holistic models of human-robot interaction.

In conclusion, emotional and social intelligence in robots is an emerging field poised to redefine human-robot interactions. By integrating sophisticated emotional and social understanding mechanisms, robots are being empowered to interact more naturally and empathetically, resonating more closely with human users. However, alongside these technological advances, careful consideration of ethical dimensions and societal impacts is essential. With continued research and development, emotional and social intelligence will undoubtedly play a foundational role in the next generation of human-centric AI systems, fostering environments where robots and humans can coexist and collaborate harmoniously.

4.3 Multimodal Communication Frameworks

In the context of embodied interaction and human-robot collaboration, the advent of multimodal communication frameworks marks a significant stride toward creating sophisticated and nuanced interfaces that allow robots to interact with humans in a robust and context-sensitive manner. This subsection delves into the integration of various communication modalities—visual, auditory, haptic, and others—to enhance the depth and richness of human-robot interactions. The frameworks discussed herein leverage these diverse channels to facilitate seamless exchanges, ensuring that robots can interpret and respond to human cues in a manner that is both intuitive and effective.

One of the primary challenges addressed by multimodal communication frameworks is the ambiguity inherent in mono-modal communication. Relying solely on a single modality, such as speech or vision, can lead to misunderstandings and misinterpretations, especially in complex environments or when dealing with nuanced interactions. Multimodal systems aim to overcome these limitations by integrating multiple input modalities, allowing robots to gather richer context and disambiguate signals when necessary.

The strength of multimodal communication lies in its ability to combine and synthesize data from disparate sources. For instance, a robot's visual sensors can capture gestural cues, while auditory sensors pick up vocal instructions, and haptic sensors can sense touch or force. These systems use advanced sensor fusion techniques to create a coherent and contextually informed model of the task environment [25]. The integration of these sensory modalities enables robots to make more informed decisions and adapt their behaviors dynamically based on the input received across various channels.

Several approaches have been explored for the development and implementation of multimodal communication frameworks. One prevalent method involves the use of probabilistic models that can handle uncertainty and model the combined influence of different modalities. For example, hidden Markov models (HMM) and Bayesian networks have been applied to interpret combined verbal and non-verbal cues [5]. These probabilistic techniques are effective at representing and reasoning under uncertainty, which is crucial for real-time human-robot interactions where sensory noise and ambiguities are commonplace.

Another promising approach is the use of neural networks, particularly those employing deep learning architectures. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can process high-dimensional data from multiple modalities, such as images, sounds, and text inputs, to learn complex patterns and associations. These networks are adept at identifying subtle correlations and can be trained to perform tasks such as speech recognition, gesture detection, and emotional understanding [48].

The development of multimodal frameworks also incorporates the concept of embodiment, wherein robots are designed to perceive and act in the real world in ways that mimic human interactions. This involves the creation of anthropomorphic features and behaviors that align with human social expectations, facilitating a natural and seamless interaction [3]. By focusing on embodied intelligence, robots can interact more fluently within human environments, effectively translating multimodal inputs into actions that are contextually appropriate and socially acceptable.

However, the creation of effective multimodal communication frameworks is not without its challenges. One significant limitation is the computational cost associated with processing and integrating high-dimensional data from multiple sources. Real-time performance is often essential, especially in dynamic and unpredictable environments, necessitating the development of efficient algorithms and models that can operate within the constraints of limited processing power and bandwidth [21]. Advances in neuromorphic computing and other efficient hardware architectures are promising solutions to alleviate these computational burdens.

Moreover, there is the challenge of synchronizing data from disparate modalities which may have different update rates and latency characteristics. Effective synchronization is crucial to maintaining temporal coherence and ensuring that robot actions are both timely and contextually relevant [68]. Models such as Kalman filters and temporal convolutional networks (TCNs) are often employed to address these issues, providing a means to align and integrate sensory inputs within a unified temporal framework.

Emerging trends in the field point toward increasingly sophisticated models that are capable of contextually interpreting human-robot interactions. The integration of AI-driven natural language processing (NLP) systems with contextual understanding allows robots to engage in more meaningful dialogues, understanding the nuances of human language and adjusting their responses based on the context. Emotion recognition technologies, leveraging both visual and auditory data, further enable robots to detect and appropriately respond to human emotions, enriching the overall interaction experience.

In conclusion, multimodal communication frameworks represent a transformative shift toward more intuitive and effective human-robot interactions. By integrating and synthesizing data across multiple channels, these systems offer robust solutions to interpret complex human inputs in varied environments. Future research is set to explore further integration of machine learning algorithms for adaptive learning, as well as the development of more computationally efficient models to manage the significant data processing demands inherent in multimodal systems. The ongoing evolution of these frameworks holds great promise for enhancing the capabilities of embodied AI, thereby fostering more symbiotic partnerships between humans and robots. Through continued innovation, these systems will not only advance the field of robotics but also enhance the quality of interactions between humans and machines, paving the way for deeper integration of robots into everyday life.

4.4 Collaborative Task Performance

The subsection on collaborative task performance within embodied interaction and human-robot collaboration explores essential frameworks and methodologies needed to facilitate seamless human-robot cooperation. By achieving shared goals and synchronized actions, these frameworks improve both the quality and efficiency of task execution, making them integral to the successful integration of robots into dynamic environments across industrial and daily life scenarios. Understanding these collaborative frameworks is crucial for maximizing the potential of human-robot teams.

A key aspect of collaborative task performance is the strategic allocation and coordination of roles between human and robot teammates. Methodologies for dynamic role allocation have been developed, each offering distinct strengths and limitations. Dynamic role assignment, for instance, involves adaptively distributing tasks based on real-time assessments of both human and robot capabilities along with environmental factors. Flexibility in role distribution is vital in complex environments where unforeseen events can impact task execution. Cognitive architectures in robots are instrumental in facilitating efficient task division by predicting future states and adjusting roles to mitigate risks and optimize outcomes [26].

Interactive learning strategies play a pivotal role in enabling robots to refine their collaborative skills over time. By observing and adapting to human behavior, robots become more adept partners [55]. Techniques like imitation learning equip robots with the ability to acquire new skills through the emulation of human actions. Concurrently, reinforcement learning enables them to practice and polish these skills across different scenarios, thus enhancing their adaptability and efficiency in collaborative settings [30]. Adaptive learning is further enhanced by continuous feedback loops that refine and reinforce desired behaviors.

Real-time feedback and control systems are critical to effective human-robot task collaboration, enabling instantaneous adjustments during activities to ensure harmonious alignment of human and robotic actions [69]. The design challenge involves creating feedback mechanisms that are intuitive for human operators while being robust enough to quickly process and respond to diverse sensor inputs. Advanced control theories are increasingly employed to sync actions, optimizing physical and communicative interactions for improved task outcomes [70].

The integration of emotional and social intelligence into collaborative frameworks is a burgeoning interest area. Such intelligence allows robots to better interpret and respond to human social cues, enhancing trust and comfort in human-robot collaborations [54]. Emotional intelligence capabilities notably improve interaction quality, particularly in tasks requiring an understanding of human emotional states. Moreover, emotional and social intelligence makes environments shared with robots less anxiety-inducing for humans, supporting more efficient and acceptable interaction modalities [71].

Currently, there is a trend towards more symbiotic collaboration forms where humans and robots equally contribute to tasks rather than one assisting the other. This shift is driven by advancements in autonomous decision-making capabilities and neuromorphic computing, enabling robots to undertake complex tasks with minimal oversight [29]. The new generation of collaborative robots, or 'cobots', is designed for seamless team integration, offering suggestions and making decisions that boost overall team performance [72].

Despite significant progress, challenges persist. Ensuring collaborative task safety and robustness in uncertain environments requires sophisticated models to accurately predict collaborative action outcomes and anticipate potential failures [73]. Additionally, maintaining and enhancing the trustworthiness of collaborative robots is essential, as trust heavily influences the acceptance and effectiveness of robotic systems in human-majority settings [74].

Future research in collaborative task performance will likely concentrate on refining algorithms for role flexibility and adaptive learning, integrating advanced emotional intelligence frameworks, and developing predictive models to ensure task success across varying conditions. As robots evolve, improving their natural and intuitive engagement abilities will be crucial for fully realizing their potential in collaborative settings [29].

In conclusion, the exploration of collaborative task performance in human-robot interactions holds great promise for advancing the contributions of robots to human endeavors. Developing sophisticated collaboration strategies that encompass dynamic role allocation, interactive learning, and emotional intelligence brings us closer to achieving seamless and efficient human-robot synergies capable of transforming industries and enhancing daily life experiences.

4.5 Evaluation and Metrics for Human-Robot Collaboration

In evaluating human-robot collaboration (HRC), a multifaceted approach is essential, one that encompasses both quantitative and qualitative measures to thoroughly assess interaction effectiveness and quality. This subsection offers a comprehensive overview of the current methodologies employed in evaluating HRC, discusses their advantages and limitations, and anticipates future directions in this evolving research domain.

Effective evaluation of HRC necessitates a blend of usability, user experience, and task performance metrics. Usability metrics typically involve assessing the ease of use, efficiency, and intuitiveness of interaction interfaces. As depicted in the study of embodied conversational agents [75], subjective usability surveys provide valuable insights into the user's interaction experiences. These surveys often encompass questions related to the system's ease of learning, efficiency, memorability, error frequency, and user satisfaction. The integration of these usability metrics with subjective evaluations, such as the System Usability Scale (SUS), can offer a richer understanding of user experiences.

User experience metrics focus on the emotional, cognitive, and behavioral responses elicited by the HRC system. These metrics become particularly pertinent when dealing with socially interactive robots that embody empathetic responses to human social cues [56]. Empathy recognition systems, as seen with Zara the Supergirl, assess emotional exchanges between humans and robots through advanced affective computing techniques. Such measures often involve sentiment analysis and physiological responses, providing a rich dataset for evaluating user emotional engagement and overall experience.

Task performance metrics emphasize the collaborative outcome of human-robot teams. These quantitative metrics can include measures such as task completion time, accuracy, error rates, and efficiency. The successful implementation of joint tracking within interactive settings, such as in the BEHAVE dataset [76], highlights the importance of evaluating how well humans and robots can jointly accomplish complex tasks. In environments that require precise manipulation and coordination, such as healthcare or industrial settings, these metrics become critical indicators of system effectiveness.

In discussing task performance, dynamic role allocation plays a vital role, where the assignment of roles between human and robot partners fluctuates based on situational demands and capabilities. This is evident in frameworks that allow robots to adaptively modify their roles within collaborative tasks. Interactive learning mechanisms, which permit robots to improve their collaborative skills through feedback and experience, highlight the evolving nature of HRC. These systems are often evaluated using metrics of adaptability, learning rate, and the quality of learned strategies, reflecting their intrinsic ability to enhance cooperation over time.

To support the objective assessment of human-robot interactions, various evaluation frameworks have emerged, integrating multi-modal data collection. The use of high-fidelity simulations allows for controlled experimentation and enables the testing of interactions in replicated settings. These simulators are integral to preemptively identifying potential pitfalls in HRC before real-world deployment. However, these simulations must be complemented by sim-to-real transfer methods, ensuring that trained agent behaviors transition smoothly to real environments without loss of performance fidelity.

Emerging trends in the field emphasize the need for standardized evaluation methodologies. As current practices often vary widely across studies, standardization efforts are crucial for the effective benchmarking of HRC systems. The use of shared datasets, such as the PHASE dataset [60], enables the cross-comparison of different approaches under consistent conditions, fostering collaborative progress within the community.

Despite advances, several challenges persist in the evaluation of HRC. Capturing the nuance of human-like social cues and integrating them into robotic cognition, as highlighted by the challenges addressed in empathetic robots for older adults [27], requires refined emotional intelligence algorithms responsive to diverse human behaviors. Furthermore, ensuring that evaluations are comprehensive and contextually sensitive to real-world applications demands ongoing research and development efforts.

The future of HRC evaluation lies in integrating artificial emotional intelligence with methods that ensure safety, accountability, and transparency in interactions. There is potential for leveraging multimodal datasets and AI technologies to develop more adaptive, context-aware robotic systems. Such systems would be capable of responding proactively to human needs, cementing their role as collaborative partners across varied domains.

In summary, evaluating HRC necessitates a nuanced, multifaceted approach that blends quantitative performance measures with qualitative user experience assessments. While strides have been made toward achieving this balance, ongoing standardization efforts and integration of advanced affective and analytical technologies will be pivotal in addressing current challenges. By honing such methodologies, the research community can propel the field towards achieving more seamless, effective human-robot collaborations in future applications.

5 Simulation and Real-World Deployment

5.1 Role of Simulation in Development

In the dynamic landscape of Embodied Artificial Intelligence (AI), simulation has become an indispensable tool for development, offering unmatched safety, efficiency, and versatility. This subsection delves into the complexities and nuances of utilizing high-fidelity simulators in training embodied AI systems, examining their crucial role in advancing this burgeoning field.

High-fidelity simulation environments have emerged as critical platforms for developing embodied AI systems, providing a controlled yet comprehensive space to iterate on algorithms and architectures. These environments allow researchers to model complex interactions between agents and their surroundings, ensuring that training scenarios closely approximate the challenges and variability found in real-world settings. Notably, platforms such as Habitat [7] are lauded for their ability to render photorealistic 3D simulations, fostering stronger agent perception and navigation capabilities.

The use of simulation in Embodied AI is lauded for several strengths. Primarily, simulators negate the inherent risks associated with physical training, such as damage to hardware or environments, by encapsulating experimentation within virtual boundaries [42]. This safety is further compounded by the recent strides in procedural content generation, which enriches the learning experiences of AI agents by exposing them to diverse and unpredictable scenarios. As simulators continue to evolve, their graphical fidelity, physics simulation, and sensor integration have improved, narrowing the reality gap and making it easier to transfer learned skills to physical robots [10].

Despite these advancements, there are notable challenges and trade-offs inherent in simulation-based development. One critical limitation is the reality gap—the discrepancy between simulated and real-world environments [10]. This gap often results in suboptimal performance of AI systems when transferred to real-world contexts, necessitating additional methods like domain randomization to ensure robustness and adaptability [6]. Furthermore, simulators, despite their advancements, may still fall short in reproducing the full spectrum of physical dynamics, potentially limiting the fidelity of learned behaviors.

Interactive and human-in-the-loop simulators have emerged in recent years to counter some of these limitations by incorporating human feedback into the training process [62]. This approach aids in refining AI behaviors, leveraging human intuition and adaptiveness to guide system development. For instance, the incorporation of human demonstrations in the training loop can accelerate learning, allowing AI systems to adaptively align with human counterparts in collaborative tasks like surgery or manufacturing [42].

Simulation platforms such as EmbodiedScan [32] are notable for their contribution to multimodal perception and interaction, enabling agents to navigate and comprehend 3D environments using integrated sensory data. Such platforms highlight the importance of converging visual, auditory, and tactile inputs to enhance the responsiveness and accuracy of embodied AI frameworks.

Looking towards the future, several trends and challenges loom over the horizon for simulation in embodied AI. The integration of Virtual Reality (VR) and Augmented Reality (AR) within training environments promises to heighten realism and user immersion, potentially leading to more intuitive interaction paradigms. Furthermore, the intersection of AI and advanced simulation techniques could usher in a new era of intelligent design, allowing for more complex scenarios without the requirement of high computational costs.

Moreover, ethical considerations, such as ensuring unbiased and safe AI deployment in diverse societal contexts, underscore the need for stringent evaluation frameworks for simulation systems. As simulation increasingly becomes a cornerstone in the development pipeline, its ethical integration and oversight will be pivotal to maintaining public trust and ensuring responsible innovation.

In summary, high-fidelity simulators represent a cornerstone of embodied AI development, providing essential tools for safe, efficient, and versatile agent training. The growing diversity and functionality of these platforms underscore the field's dynamic nature, while their limitations and emerging solutions reflect the ongoing challenges faced by researchers. As simulation environments continue to evolve, they will undoubtedly play a pivotal role in shaping the future of embodied AI systems, driving its trajectory towards real-world applicability and societal integration. Emerging trends in VR, AR, and multimodal interaction will further redefine the landscape, offering exciting avenues for innovation and exploration. Consequently, researchers and practitioners alike must remain vigilant, balancing the technical promises of simulators with ethical considerations to propel the field of embodied AI forward.

5.2 Sim-to-Real Transfer Challenges

Sim-to-real transfer is a critical aspect of embodied artificial intelligence (AI), focusing on overcoming the challenges of transitioning from simulated environments to real-world deployments. While simulation offers advantages like cost-effectiveness, risk reduction, and ease of experimentation, shifting trained models from virtual settings to tangible applications presents significant hurdles. These discrepancies, collectively known as the reality gap, stem from differences in physics, sensor noise, and environmental variation. This subsection explores these barriers, conducts a comparative analysis of existing methodologies, considers potential breakthroughs, and discusses the future trajectory of research.

The reality gap poses a fundamental challenge to transfer learning due to the inherent discrepancies between controlled simulation environments and the intricate complexities of the physical world. One primary cause is the variance in physical dynamics; simulations often use simplified physics engines that fail to replicate the nuanced interplay of forces observed in reality [5]. Bridging these discrepancies requires a multifaceted approach that enhances both simulation fidelity and real-world adaptability.

Domain randomization emerges as a promising technique to mitigate the reality gap by intentionally varying simulation parameters such as lighting, texture, sensor noise, and environmental variables [43]. The core principle is to create a highly diverse training dataset within the simulator that encapsulates a wide range of potential real-world variations. While this approach enhances the robustness and generalizability of artificial agents, it can be computationally costly and may risk overfitting to simplistic noise patterns that do not capture the true complexity of reality.

Online adaptation and real-time fine-tuning are pivotal strategies for enhancing post-deployment performance. By integrating continuous learning mechanisms and feedback loops, embodied systems can adapt their learned models based on real-world observations [2]. Real-time data streams from agent interactions with the environment inform minute adjustments, improving the alignment between anticipated and actual outcomes. However, these adaptations increase system complexity and require a balance between computational efficiency and model flexibility.

Transfer learning methods have shown potential in accelerating sim-to-real transitions by leveraging pre-trained models developed in simulation for real-world applications. This approach exploits existing knowledge, facilitating quicker learning and adaptation phases [44]. Despite its promise, transfer learning necessitates refining to inherit the beneficial aspects of prior models while avoiding biases and errors intrinsic to the simulation phase.

Physics-informed AI models present a sophisticated route for enhancing sim-to-real transfer. These models incorporate advanced physical principles into the learning process, aligning simulation dynamics more closely with real-world phenomena [77]. By embedding a granular understanding of physics, these models better equip agents to handle real-world misalignments. However, this approach can increase model complexity and demand more computational resources during training and deployment.

Residual policy learning offers an innovative technique to address sim-to-real discrepancies. Here, agents adopt a dual-strategy approach by combining simulation-trained policies with real-world corrective feedback for dynamic action adjustments [46]. This complementary strategy addresses limitations inherent in each process individually, offering robust behavioral adaptations. Ensuring these dual layers function synergistically rather than antagonistically requires meticulous calibration and oversight.

The integration of high-fidelity simulations and virtual reality (VR) systems provides a rich training platform for embodied agents [78]. High-resolution graphics and accurate physics enable more realistic training scenarios, potentially yielding better generalization to real-world conditions. However, the fidelity of current simulators is not yet on par with reality, necessitating ongoing refinement. Incorporating sensor-specific calibration and noise models in the simulator can aid agents in handling post-deployment perceptual discrepancies.

Future directions in sim-to-real transfer increasingly align with embodied cognition approaches prioritizing situatedness and real-time adaptability. This involves a broader understanding of cognitive architectures that mimic human adaptation capabilities [3]. Moreover, an interdisciplinary approach drawing from neuroscience, cognitive science, and advanced machine learning is paving the way for sophisticated models that transition seamlessly between simulation and real-world settings.

In synthesis, sim-to-real transfer is an evolving field driven by the demand for more versatile, robust, and adaptive embodied AI systems. Continued research and innovation to overcome the reality gap hold the potential to unlock unprecedented applications across diverse domains, from autonomous navigation to human-robot interaction. The amalgamation of randomized domain strategies, physics-informed learning, and adaptive control frameworks represents the fulcrum for future advancements, ultimately narrowing the divide between virtual and real-world performance.

5.3 Techniques for Enhancing Sim-to-Real Transfer

The challenge of transferring knowledge and skills acquired in simulated environments to real-world applications—referred to as the sim-to-real transfer—remains a pivotal hurdle in embodied artificial intelligence (AI). This subsection delves into advanced methodologies aimed at bridging the sim-to-real gap, vital for deploying simulations into functional real-world tasks. We explore cross-domain learning, physics-informed modeling, and adaptive frameworks that enhance the effectiveness of sim-to-real transfers.

At the core of sim-to-real transfer is the concept of transfer learning, wherein models trained in a simulated environment are adapted for real-world operations. This approach leverages the vast data and rapid experimentation possible in simulations to jump-start learning in settings where real-world data is scarce or expensive to acquire [68]. A fundamental method here is domain randomization, where the simulated environment's parameters are randomized during training to expose the AI to a variety of conditions, hence increasing its robustness upon deployment in reality [21]. While domain randomization has shown promise, its success largely depends on capturing the critical variability and dynamics that the real world may present.

Physics-informed models represent another strategic avenue. By embedding accurate physical laws and constraints into simulation models, these methods strive to mimic real-world dynamics more closely, thereby reducing the reality gap. In this context, the integration of advanced physics models has been pivotal. Techniques such as residual force control synergize physics-based policies with simulation-trained policies, allowing agents to compensate for model inaccuracies by learning corrective actions in a real-world setting [79]. This approach, by aligning the simulated agent's dynamics with the real-world context, significantly enhances robustness.

Further enhancing sim-to-real transfer is the concept of morphological computation, which takes into account the physical embodiment of the agent during learning [25]. The embodiment allows the physical properties of robots to naturally perform some computations, potentially reducing the computational load on the controller and harnessing inherent stability and adaptability of the physical form. This aspect is crucial, especially when dealing with soft robotics and complex morphologies [80].

Residual policy learning provides another innovative solution for effective sim-to-real transfer. By coupling offline-trained simulation policies with real-world correction policies, it becomes feasible to dynamically adjust operations in real-time. This technique has demonstrated success in complex humanoid tasks, facilitating smooth and secure transitions from virtual into physical realms [81]. This paradigm effectively merges pre-trained behavioral patterns with adaptive policies that correct possible deviations caused by real-world dynamics.

Cross-domain adaptation, particularly via meta-learning, emerges as a promising strategy. This involves creating models that are not only robust to domain shifts but also capable of quick adaptation to new tasks or environments with limited data [82]. The framework of meta-learning provides an overarching strategy wherein agents are trained on a distribution of tasks to enable rapid learning of new tasks by leveraging prior experiences. Thus, embodied AI systems can become more proficient in real-world scenarios that deviate slightly from their simulated training conditions.

Emerging techniques also explore integrating neuromorphic computing with embodied systems, promising a leap forward in how real-time sensory inputs are processed and acted upon. Neuromorphic processors mimic human brain architectures, potentially leading to more efficient and adaptive systems that can handle complex sensorimotor tasks directly in dynamic environments [24].

Despite these advancements, significant challenges persist. A predominant issue is the discrepancy between simulated perceptions and real-world sensory feedback, often leading to inefficiencies in transferring learned behaviors. Innovations in perception models, wherein agents are equipped with advanced sensory integration frameworks, have started to mitigate this by providing richer, more contextually relevant data to the controllers [49].

In conclusion, while substantial progress has been made in sim-to-real transfer, ongoing research reveals the need for more comprehensive methodologies that blend physical realism with adaptive learning paradigms. Future directions might focus on refining multi-fidelity simulation frameworks that allow seamless exploration across varying degrees of realism, thereby nurturing a deeper understanding and manipulation of the embodiment-cognition nexus. Furthermore, collaborative efforts that integrate advancements in sensor fabrications, computational models, and AI learning frameworks are likely to chart new pathways in ensuring seamless deployment of simulation-trained embodied AI systems in the unpredictable tapestry of the real world.

5.4 Case Studies of Real-World Applications

In examining the impact of embodied AI systems in real-world applications, we not only highlight their empirical successes but also unravel the indispensable role of simulations in their evolution. This subsection navigates through diverse fields—healthcare, industrial operations, and autonomous navigation—focusing on how simulation-based development has shaped their deployment and associated challenges. By integrating academic insights alongside empirical evaluations, we aim to articulate a comprehensive narrative of embodied AI's burgeoning influence and the lessons learned along the way.

Embodied AI systems in healthcare, particularly robotic surgical systems, have shown significant potential in enhancing surgical precision and patient care. These systems are designed to mimic human hand movements, enabling minimally invasive procedures with high accuracy and reduced recovery time. Simulations play a pivotal role in their development, offering a safe environment for training and testing robotic movements under various hypothetical patient scenarios. High-fidelity simulators, for instance, allow robots to learn tasks such as suturing and organ dissection without real-world consequences, encouraging surgical innovation in a risk-free setting.

A critical case study involved a simulated robotic assistant designed for complex orthopedic surgeries, trained under a vast array of procedurally generated scenarios to mimic real surgical challenges. Although these systems have demonstrated potential to increase surgical efficiency and accuracy, they face challenges in adapting AI decisions to the dynamic nature of human physiology and unpredictable surgical complications. The reality gap, where simulated models fail to capture real biological variability, remains a critical hurdle [31].

In manufacturing and logistics, the transformation brought about by embodied AI systems via automation and enhanced human-robot collaboration is noteworthy. These robots are often trained within simulated environments for tasks like assembly, packaging, and material handling before deployment. Simulation platforms model complex industrial processes, refining robots' decision-making for optimized real-world performance [74].

A prominent application is collaborative robots, or cobots, which work alongside humans on production lines. Through simulation, these cobots perfect their movement paths and interaction protocols to minimize safety risks and enhance synergy with human workers. However, smoothly integrating such systems into existing workflows poses challenges, particularly in ensuring precise synchronization with human activities and transferring simulation-trained behaviors to dynamic real-world conditions [29].

Autonomous vehicles (AVs) and drones exemplify the synthesis of embodied AI with navigation challenges. Simulations create digital replicas of environments complete with varying conditions to test AV navigation algorithms rigorously. This calibration of sensory inputs and real-time decision-making capabilities reduces the risks of real-world deployment.

Simulation-driven AV development provides insights into environmental data interpretation and hazard anticipation. Sophisticated simulators explore scenarios where AVs navigate mixed-traffic environments, responding to dynamic obstacles and pedestrian movements [83]. The reality gap, however, poses discrepancies between virtual success and real-life accuracy due to sensor inaccuracies and unforeseen road conditions [31].

The trajectory of embodied AI in real-world applications underscores the necessity of advancing simulation fidelity and refining sim-to-real transfer methodologies. Developing physics-informed simulations could bridge the reality gap, offering reliable transitions of AI systems from virtual environments to the unpredictability of real-world contexts. Future research should consider human-in-the-loop simulations to enhance AI adaptability and resilience through real-time human feedback [84].

Looking forward, the success of embodied AI systems will rely on a symbiotic relationship between simulation environments and real applications, where iterative feedback loops enhance capabilities. By addressing emerging challenges, embodied AI systems are poised to achieve unprecedented integration and efficiency, heralding a new era of human-AI collaboration across diverse societal and industrial landscapes.

5.5 Current Trends and Future Directions in Simulation

The realm of simulation for embodied artificial intelligence (AI) is undergoing transformative progress, focusing on creating more sophisticated, robust, and versatile models that can seamlessly integrate into real-world applications. Embodied AI necessitates advanced simulation environments that closely mimic real-world conditions to bridge the so-called sim-to-real gap—a major frontier in the field. This subsection explores the current trajectories and future paths shaping simulation methodologies for embodied AI, addressing the challenges and opportunities presented by emerging technologies and research.

A primary trend in simulation is the incorporation of Virtual Reality (VR) and Augmented Reality (AR) to create immersive training environments. These technologies enable embodied agents to engage with richly detailed and interactive scenarios that enhance behavioral fidelity during training. VR and AR facilitate a more tangible sensory experience, providing higher precision feedback mechanisms that contribute to realistic agent responses. This integration encourages the development of simulation platforms that align closely with real-world dynamics, enhancing agents' ability to generalize learned behavior from virtual settings to physical environments [16].

Scalability and distributed simulation are also gaining traction as pivotal components of future simulation frameworks. Large-scale, distributed environments can simulate interactions at a massive scale, enabling the testing and optimization of multi-agent systems in diverse virtual milieus. This approach mirrors real-world complexities more closely, offering a comprehensive testbed for agents meant to operate within interconnected settings—a feature particularly pertinent for training agents in autonomous navigation and collaborative robotics [61].

Another prominent challenge in simulation research is ethical considerations and impact assessment. As simulation-trained embodied AIs increasingly partake in societal and industrial applications, the need to address potential biases, unfair outcomes, and data privacy concerns becomes critical. Ensuring that simulation environments are ethically sound and that the models interpret situational data responsibly is essential. This challenge underscores the importance of establishing frameworks that can forecast and mitigate potential societal impacts prior to real-world deployment [8].

In evaluating these trends, it's crucial to consider the breadth of methodologies aiming to facilitate effective sim-to-real transfers. Techniques such as domain randomization—where simulation parameters are varied extensively to train agents for adaptability—are instrumental in closing the reality gap. These paradigms force agents to learn robustly by confronting them with numerous environmental permutations, ensuring that when encountering unforeseen real-world conditions, the trained models can retain reliable performance. Advances in physics-informed simulations also aid this transition by incorporating nuanced physics-based modeling, allowing simulations to approximate the real-world physical interactions more accurately than before [60].

In terms of future directions, the potential impact of multimodal interactions within simulations—as driven by advanced multimodal large language models (MLLMs)—cannot be overstated. Such models promise to revolutionize how embodied AI systems interpret mixed sensory inputs, enhancing situational awareness and response efficacy. With further integration of these models into simulation environments, virtual agents will benefit from a holistic understanding grounded in textual, auditory, and visual data [36].

Moreover, to address the demands of ongoing research, it is clear that broader collaborative efforts are necessary. Cross-disciplinary collaborations, blending insights from cognitive science, machine learning, and robotics, will drive innovations in developing more effective and ethically responsible simulation environments. These partnerships must extend into industry to ensure applied research translates into impactful products and services that align with both technological and societal goals [61].

Ultimately, the future of simulation in embodied AI resides in creating environments that not only empower agents to perform effectively in controlled settings but that also equip them with the versatile problem-solving skills required in uncontrolled, dynamic real-world scenarios. As researchers push the boundaries of what simulation can achieve, attention must remain on creating frameworks that emphasize robust, ethical, and interdisciplinary models to sustain the evolving landscape of embodied AI. These endeavors are pivotal to harnessing the full potential of embodied AI, ensuring that it continues to contribute meaningfully to society's technological fabric.

6 Ethical, Societal, and Evaluation Considerations

6.1 Ethical Frameworks and Implications

The deployment of Embodied AI in real-world scenarios brings forth numerous ethical considerations, central to which is the development and integration of robust ethical frameworks. These frameworks are fundamentally designed to guide the creation, deployment, and management of embodied AI systems, ensuring they align with societal values and ethical principles. This subsection explores various ethical paradigms, analyzes their strengths and weaknesses, and discusses the critical implications of adopting these frameworks in the context of embodied AI.

At the intersection of technology and ethics, the endeavor to define ethical frameworks involves adapting long-standing philosophical theories such as deontology, utilitarianism, and virtue ethics to the unique challenges posed by embodied AI. Deontological ethics focuses on the adherence to rules and duties, which can translate into the development of stringent regulatory policies and design guidelines for AI systems. In contrast, a utilitarian approach prescribes the maximization of overall happiness, pushing for AI outcomes that generate the greatest benefit for the greatest number of people [41]. Virtue ethics encourages the cultivation of moral character, implying that designers and developers should embody virtues such as fairness and responsibility throughout the AI development lifecycle.

Empirical research underscores the necessity of such ethical frameworks by highlighting potential risks associated with embodied AI. Systems that lack ethical grounding may inadvertently perpetuate biases or invade privacy, leading to eroded public trust [15]. To mitigate these risks, standards for ethical design have been proposed, focusing on principles such as fairness, transparency, and accountability. Explainability, a prominent feature of responsible AI design, demands that systems provide clear, understandable rationales for their decisions, fostering accountability and promoting user trust [11].

Nevertheless, ethical frameworks are not without their trade-offs and challenges. An overemphasis on rigid deontological guidelines, for instance, may stifle innovation, whereas a utilitarian focus can lead to ethically questionable cost-benefit analyses where minority rights are compromised for utilitarian gains. Furthermore, embedding ethical principles within AI systems is a non-trivial task, given the complexity and contextual variability of AI interactions [40].

Addressing ethical dilemmas in embodied AI also requires consideration of data privacy and usage. As these systems increasingly rely on large volumes of sensory and contextual data, the potential for misuse or breach becomes a pressing concern [38]. Solutions to these issues involve implementing data anonymization techniques, securing data through encryption, and incorporating user consent mechanisms to bolster privacy measures.

One emerging trend is the integration of explainable AI (XAI) methods to address ethical concerns. By providing transparency into the decision-making processes of AI systems, XAI fosters trust and empowers users with the ability to scrutinize AI actions. For example, residual policy learning techniques can enhance operational understanding by combining simulation-trained policies with corrective real-world actions [85]. Additionally, innovative human-in-the-loop approaches ensure that embodied AI systems remain aligned with ethical standards even after deployment, thus continually validating their ethical and safe functionality [64].

Significant strides are being made in developing comprehensive ethical guidelines specifically tailored for embodied AI, as evidenced by initiatives such as Ethical AI by Design. These incorporations into AI frameworks suggest that ethical considerations should be integrated from the onset of system design, rather than applied as retroactive corrections [86]. Such proactive encompassing strategies pave the way for systems that are inherently ethical and socially beneficial.

In conclusion, the ethical implications of deploying embodied AI underscore the importance of multi-faceted ethical frameworks that consider the full spectrum of philosophical, legal, and practical challenges. Future research directions could explore frameworks that weigh contextual factors more dynamically, allowing for flexible yet consistent ethical oversight across diverse applications. Furthermore, collaborative cross-disciplinary efforts are essential to address these evolving challenges, offering mechanisms for incorporating societal, scientific, and ethical insights into the design of new generations of embodied AI. By synthesizing these and related insights, we can move toward an ethically sound future where embodied AI is aligned with human values and societal benefit.

6.2 Societal Impacts and Human-AI Interaction

Embodied Artificial Intelligence (AI) represents a paradigm shift with profound impacts on societal norms, human interaction dynamics, and the framework of technological acceptance. This subsection discusses the integration of Embodied AI systems into daily life, emphasizing how these systems are reshaping human-AI interactions, influencing societal structures, and setting the trajectory for future developments.

The fusion of physical and cognitive capabilities within embodied AI systems creates a multifaceted landscape of societal impacts. The deployment of such systems—like social robots in healthcare—introduces new methods of human-robot interaction, prompting a reevaluation of social norms and ethical considerations [2]. These systems transcend their role as mere tools to become active social participants, capable of interacting, learning, and adapting within human environments.

A crucial element in societal engagement with embodied AI is the integration and acceptance of these systems. Acceptance levels largely depend on factors like perceived utility, trust, and cultural receptivity [3]. Successful integration is marked by enhancements in societal functions, such as improved healthcare delivery or elder care support, whereas negative perceptions may lead to resistance or anxiety. Empirical studies on embodied language acquisition highlight the potential for improved interaction paradigms through advanced speech recognition and social cue interpretation, fostering greater acceptance of robots in social contexts [66].

Further, embodied AI systems are transforming social norms and expectations. These systems, capable of executing daily tasks, interacting emotionally, and making decisions, challenge traditional human labor roles and interpersonal interactions. As AI agents manage more decision-making processes previously handled by humans, perceptions of agency and human engagement are shifting. Symbol emergence in robotics exemplifies this change, enabling AI to develop its interpretations and symbolic language grounded in human social interaction [4].

One critical aspect redefining human-AI interaction paradigms is the empathetic potential of embodied systems. By incorporating emotion recognition and response mechanisms, AI systems can offer personalized, responsive user experiences, leading to increased user satisfaction and social engagement. Despite these advances, risks include potential over-reliance or emotional attachment to devices, possibly affecting human psychological and social development adversely [45].

The ongoing evolution of human-agent interactions fuels technological innovation while underscoring ethical concerns related to privacy, security, and autonomy. Embodied AI's ability to gather contextual and sensory information necessitates robust data protection mechanisms and transparent, accountable AI operations. Trust is foundational in this dynamic, requiring AI systems to demonstrate reliability and ethical compliance to enhance societal acceptance.

The technological and educational sectors play a crucial role in shaping the future landscape for societal acceptance of embodied AI. Public awareness campaigns, educational initiatives, and inclusive design practices can mitigate resistance and dispel misconceptions through informed understanding and engagement. Additionally, interdisciplinary frameworks combining insights from psychology, sociology, and AI can foster a holistic approach to acceptable AI integration [87].

In conclusion, embodied AI stands ready to redefine societal norms and interaction paradigms through its transformative capabilities in human-AI integration. The field's trajectory will largely depend on the adaptability of social systems, the robustness of regulatory frameworks, and the progressive alignment of embodiments with human-centered ideals [88]. Future research must address the nuanced challenges presented by these systems—from existential and ethical dilemmas to technological constraints—ensuring that embodied AI contributes positively to societal advancement without undermining human agency and societal cohesiveness.

6.3 Privacy, Security, and Trust Considerations

In the realm of Embodied Artificial Intelligence (Embodied AI), privacy, security, and trust considerations present a significant frontier that must be navigated with both diligence and sophistication. This subsection explores the complex interplay of privacy protection measures, security enhancement strategies, and trust-building mechanisms within embodied AI systems, scrutinizing their current implementations, potential challenges, and future directions.

At the core of privacy concerns in embodied AI is the need to handle extensive user data responsibly. Embodied systems often rely on continuous data streams from sensors capturing audio, video, biometrics, and environmental information. Privacy protection begins with data minimization principles, advocating that only essential data should be collected and retained. Techniques such as data anonymization and pseudonymization serve as foundational strategies to obscure personally identifiable information, thereby reducing risks if data breaches occur. For example, in applications like musculoskeletal humanoids [89], the system continuously gathers sensitive health data, which necessitates strict protocols to ensure data privacy while maintaining the system's utility.

Moreover, robust encryption protocols are critical in safeguarding data integrity and confidentiality. For instance, communication between the AI systems and external interfaces, such as in bimanual telemanipulation systems [90], must utilize advanced cryptographic techniques to prevent interception and unauthorized access. End-to-end encryption is one method that ensures data remains secure from the point of collection to its final destination, thus maintaining user privacy and system security.

Security concerns extend beyond data protection to encompass the physical security of embodied AI systems. Unauthorized access to these systems, whether through digital means or physical tampering, could lead to potentially harmful consequences, especially in applications involving robots operating in close proximity to humans [91]. Strategies for enhancing security include multi-factor authentication, secure access protocols, and continuous monitoring for anomalies indicative of security breaches. Additionally, the integration of residual policy learning techniques, as seen in models like [79], can provide added layers of security by enabling systems to detect and respond to unexpected behaviors in real-time.

A pivotal component in the deployment and acceptance of embodied AI systems is trust. Trust can be cultivated through transparency, where the system's decision-making processes and data usage policies are clearly communicated to users. Explainable AI (XAI) frameworks can demystify the inner workings of complex algorithms, building confidence in system reliability and decisions, as highlighted in systems using embodied flight [92]. Trust is further reinforced by consistently demonstrating system reliability through testing and benchmarking against standardized performance metrics [49].

Ethical considerations also play a crucial role in constructing trust. Embodied AI systems must adhere to ethical standards and regulations that prioritize user welfare and societal norms—a concept explored under various theoretical frameworks. Ensuring fairness, accountability, and bias mitigation in AI systems is a growing area of concern, especially as these technologies intersect with sensitive domains like healthcare and social robotics.

Emerging trends in privacy, security, and trust in embodied AI suggest a move towards integrating AI ethics with robust technical solutions. The development of decentralized AI architectures, such as federated learning models, is emerging as a promising avenue to enhance data privacy by keeping data localized. Furthermore, the convergence of AI with advanced neuromorphic hardware offers avenues for improving the efficiency, security, and trustworthiness of embodied AI systems [6].

As embodied AI systems continue to evolve, the challenges in privacy, security, and trust will likewise grow more complex. The advancements in AI pose ethical dilemmas surrounding autonomy, decision-making, and the erosion of human oversight in critical applications. There is a pressing need for interdisciplinary collaborations to create robust frameworks capable of addressing these challenges holistically. Future research must focus on developing adaptive systems that are resilient to vulnerabilities while considering the ethical implications of their deployment in diverse settings.

In conclusion, safeguarding privacy, augmenting security, and fostering trust are indispensable to realizing the full potential of embodied AI. By advancing these areas, the field can pave the way for the widespread acceptance and responsible use of these innovative technologies, ensuring they serve to enhance, rather than detract from, human society. As research continues to unfold, maintaining a balanced dialogue across disciplines will be critical to achieving comprehensive solutions that align technical innovations with ethical imperatives.

6.4 Evaluation Metrics and Methodologies

In the domain of embodied artificial intelligence (AI), evaluating systems involves intricate methodologies and a multifaceted set of metrics. This subsection explores the robust evaluation frameworks tailored for embodied AI systems, emphasizing their efficacy, resilience, and compliance with ethical norms. The evaluation of embodied systems is particularly complex due to the integration of physical embodiment with cognitive processes, necessitating diverse and comprehensive assessment tools.

A comprehensive evaluation framework for embodied AI encompasses ethical, technical, and societal dimensions, echoing the privacy, security, and trust considerations discussed previously. The first dimension often involves ethical implications focusing on the responsible deployment of AI systems in real-world settings. Ethical evaluation metrics consider factors such as privacy, fairness, and accountability, which resonate with the need for trust-building mechanisms. Standards for ethical design, as stated in [93], are integral to this process and provide guidelines to ensure that embodied systems act within acceptable moral frameworks.

Technically, the performance of an embodied AI system can be assessed by quantifying its ability to perform specific tasks with accurate feedback mechanisms in place—a continuation of discussions on technical proficiency and security measures. Among technical metrics, reliability and error rates are crucial, as these systems often operate in dynamic and unstructured environments. For instance, in measuring morphological computation, systems must determine how morphology aids in achieving a desired work output [1]. Performance assessments also involve benchmarking initiatives that use standard scenarios to provide performance baselines. These benchmarks help illustrate strengths, such as fast processing or adaptability, and elucidate weaknesses, such as limitations in generalization across different domains.

From a societal perspective, the measurement of human-agent interaction quality becomes a critical aspect, complementing the trust considerations of earlier sections. Studies reveal that the degree to which a user perceives a system as relatable or trustworthy greatly influences its acceptance [74]. Evaluating the societal impact involves metrics like user satisfaction, ease of use, and the quality of interaction between humans and robots, with frameworks such as Human-AI Interaction paradigms offering insights for cohesive engagement [94].

Different methodologies have emerged to tackle these evaluation challenges, bridging the comprehensive perspectives needed for both technical and societal assessments in future and current AI integration. Comparative analysis of different approaches highlights the trade-offs and strengths inherent in each. For example, user-centric evaluations, such as task-based performance metrics, offer an in-depth analysis of the dynamic performance across real-world settings. Usability and user experience are often derived from these methodologies and provide insights into the practical application of embodied AI systems [95]. However, these approaches may overlook critical technical performance metrics like latency or energy efficiency.

Hybrid strategies that integrate qualitative feedback with quantitative measures are gaining prominence, offering a more balanced appraisal of embodied systems. Such methodologies address the dichotomy between functional effectiveness and user satisfaction, offering a holistic view of system performance [29]. Emerging challenges within the evaluation domain include ensuring adaptability and resilience in diverse contexts, particularly as embodied AI systems evolve to operate more autonomously [73]. Evaluations must also adapt to cover the societal implications of AI systems more thoroughly, particularly as they become more pervasive in everyday life.

A significant trend in the evaluation of embodied AI is the increasing push towards standardizing metrics and methodologies, anticipating the future needs for multidisciplinary collaboration. This includes initiatives focused on creating unified performance benchmarks that facilitate cross-comparison among diverse systems, fostering innovation and improvement in the field [96]. However, achieving such standardization presents challenges due to the complexity and heterogeneity of embodied AI environments.

Technical metrics for performance and safety are also critical. This includes real-world task performance metrics, which focus on factors like accuracy and robustness. Safety metrics, on the other hand, scrutinize systems' resilience in preventing harm or failure during operation. Formal definitions of these metrics often accompany methodologies and can include modeling of risk assessments or reliability engineering criteria.

Moreover, emerging methodologies emphasize the importance of explainability and transparency in evaluation, addressing societal demands for trustworthy AI. Systems that integrate neuro-symbolic approaches, for instance, benefit from improved performance with reduced training data, demonstrating how symbolic reasoning enhances neural network-based learning by providing an additional layer of interpretability [29].

In conclusion, the evaluation of embodied AI systems demands a structured framework that balances ethical, technical, and societal considerations, employing both qualitative and quantitative assessment methods. Future evaluation trends will likely emphasize greater standardization and the integration of ethical metrics to ensure systems align with societal values. Comprehensive, multifaceted evaluation strategies will be pivotal in advancing embodied AI, ensuring these systems are robust, effective, and ethically accountable, setting the stage for the multidisciplinary discussions to follow. The field continues to evolve, with ongoing research striving to synthesize current methodologies and develop innovative frameworks that better capture the multi-dimensionality of embodied AI systems.

6.5 Multidisciplinary and Collaborative Approaches

The development and deployment of embodied AI cannot be effectively realized through singular disciplinary approaches. The complex interplay within ethical, societal, and evaluative processes necessitates the integration of diverse scientific and societal perspectives, fostering a vibrant ecosystem of multidisciplinary collaboration. This subsection explores how these cross-disciplinary engagements enhance the ethical frameworks, societal implications, and evaluation methodologies of embodied AI.

Embodied AI inherently demands a holistic understanding that draws parallels between cognitive sciences, ethics, robotics, psychology, sociology, and computer science. The interplay between these disciplines is vital to ensuring that AI systems are not only technically proficient but ethically grounded and socially acceptable. As scientists and engineers strive to align AI with human values, insights from sociology and ethics provide frameworks to navigate social norms and moral boundaries. Studies that scrutinize socially interactive robots, for instance, underscore the necessity of understanding embodiment from philosophical, psychological, and sociological perspectives, creating a richer dialogue between humans and machines [63].

A comparative analysis reveals that while certain disciplines offer robust frameworks for understanding embodied intelligence, others provide the tools for practical implementation. Cognitive science contributes foundational insights into sensorimotor integration and the mind-body problem, elucidating how embodied agents perceive and respond to stimuli [3]. Meanwhile, engineering disciplines translate these theories into tangible architectures and control systems, enabling the realization of perceptive and emotionally intelligent robots [56]. Nevertheless, the distinction often lies in the approach: psychology and cognitive science focus on the representation and processing of sensory data, while engineering prioritizes the actuation and manifestation of these processes into physical actions [97].

Collaborative frameworks bring together disparate elements of expertise, creating a synergy that is critical for advancing embodied AI. Efforts such as the PhD-level collaborative initiatives in virtual reality and AI propose scalable long-term strategies for integrating semantics and sensorimotor experiences, which are crucial for holistic AI development [16]. These collaborations capitalize on shared goals, drawing on the computational strengths of machine learning and the narrative reasoning skills of cognitive sciences to develop empathetic and socially aware AI agents [98].

However, these interdisciplinary endeavors face challenges, notably in reconciling the cognitive authenticity of AI's emotional responses with the computational constraints of real-time processing. The endeavor to translate the complex emotional spectra into computationally feasible algorithms requires concerted efforts from neuroscience, psychology, and computer science domains, as illustrated by the work on emotional recognition and empathy in AI systems [27].

As the field matures, emerging trends include the heightened focus on multi-modal interaction where AI systems integrate diverse data streams to enhance decision-making and communication with humans [99]. This integration necessitates not only technical adaptability but also ethical vigilance, particularly when interactions influence social dynamics and individual privacy. The ongoing development of empathetic AI, which responds emotionally appropriate to human cues, highlights a future trend towards more personalized and context-sensitive interactions [36].

Another trend is the use of reinforcement learning enhanced by human feedback, which brings ethical considerations into sharper focus as systems learn and adapt in real-world scenarios [100]. This necessitates robust evaluation frameworks that assess not just technical performance but also ethical alignment and societal impact, presenting an ongoing challenge for the field.

A synthesis of multidisciplinary insights reveals the profound complexity and transformative potential of embodied AI but also calls for cautious and responsible progress. As research progresses, the integration of diverse perspectives will be crucial to navigate the ethical and societal dimensions that underpin technically sophisticated AI agents. Future research can benefit from continued interdisciplinary dialogue, leveraging the methodological rigor of computer science, the ethical foresight of humanities, and the societal awareness of social sciences to build AI systems that not only complement human abilities but also align with human values.

In conclusion, the advancement of embodied AI is critically contingent on robust multidisciplinary collaboration that bridges the technical, ethical, and societal realms. This integrative approach fosters the development of AI systems that are not only technically proficient but ethically and socially responsible. As we move forward, reinforcing these collaborations with standardized evaluation metrics and participatory design processes could enrich AI's alignment with societal norms and expectations, charting a course for future developments that are inclusive, ethical, and innovative. By fostering a community of interdisciplinary practitioners, the embodied AI field can tackle the multifaceted challenges it faces, ensuring its safe, responsible, and beneficial integration into society.

7 FUTURE TRENDS AND RESEARCH DIRECTIONS

7.1 Emerging Technological Innovations

The rapid evolution of embodied artificial intelligence (AI) is poised on the precipice of transformative technological innovations that are set to greatly enhance the capabilities of embodiment in AI systems. As we delve into the nexus of emerging advancements in processing power, sensor integration, and action planning algorithms, this subsection aims to dissect these significant technological strides, their interactions, and implications for the future of embodied AI.

At the forefront of these innovations is the co-designing of spatial AI systems, which involves a holistic integration of algorithms, processors, and sensors specifically engineered to meet the demands of next-generation embodied systems. This approach opens up new possibilities for AI systems operating within geographic and spatial contexts, enhancing their ability to process and interpret complex environmental data [7]. Such innovations are integral in transitioning from traditional AI models towards systems that mimic human-like understanding and interaction with their surroundings, thereby enabling more effective autonomous operations.

Neuromorphic computing stands as another pivotal frontier, offering transformative potential in the realm of efficient, adaptive embodied systems. Neuromorphic processors, inspired by the architecture and functionality of the human brain, promise to significantly advance embodied intelligence by endowing systems with enhanced processing capabilities and energy-efficient operation [15]. By mimicking neural patterns, these processors facilitate rapid, parallel processing of multimodal sensory inputs, which is crucial for real-time task adaptation and fulfillment. This capability is particularly pertinent to the development of intelligent robotic systems that must process vast streams of data quickly, such as in dynamic and highly interactive environments.

Parallelly, the integration of large language models (LLMs) into embodied AI systems emerges as a promising avenue, enhancing the systems' ability to interpret, plan, and execute complex tasks through improved language processing capabilities [101]. The convergence of LLMs with embodied AI offers a multifaceted approach to understanding and interacting with environments using natural language, thereby bridging the gap between human and machine communication. LLMs, when incorporated into embodied systems, empower them with the ability to parse intricate language commands and conduct tasks that require nuanced contextual understanding, promising a leap forward in achieving autonomous, intelligent behavior in unstructured settings.

Despite these advancements, challenges remain, particularly concerning the sim-to-real transfer learning capabilities of embodied AI systems. The reality gap, defined as discrepancies between models trained in simulation and their real-world performance, continues to be a significant hurdle. Overcoming this gap requires advances in domain randomization techniques, which involve diversifying simulation parameters to enhance the robustness of AI agents when they transition from virtual to real environments [10]. By refining these methodologies, researchers can improve the generalizability and reliability of embodied AI systems, ensuring they perform effectively across diverse real-world scenarios.

Another critical exploration area is the co-optimization of morphology and control [6]. This process aims to simultaneously evolve both a robot's physical structure and its control strategies to maximize efficiency and performance. The inherent complexity in maintaining balance between morphological adaptations and effective control mechanisms poses challenges, yet it is essential for realizing more adept and adaptable embodied systems. Through innovative approaches like morphological innovation protection, which allows temporary relaxation of optimization pressures to accommodate morphological changes, it becomes feasible to reach high levels of systemic harmony and functionality.

Looking towards the future, the potential integration of these technological innovations heralds a paradigm shift in embodied AI. The combination of neuromorphic hardware, advanced language models, and co-optimization strategies could lead to the development of sophisticated systems capable of handling open-ended tasks autonomously, adapting to novel environments with minimal supervision. This forward trajectory also invites further interdisciplinary collaborations aiming to synthesize insights from cognitive sciences, neuroscience, and AI, thereby fostering the creation of more human-like agents that reflect an advanced understanding of both biological and artificial intelligence [38].

In conclusion, the burgeoning field of embodied AI stands on the cusp of remarkable technological innovations. As researchers and industry experts continue to explore and integrate these cutting-edge developments, the potential for enhanced real-world applications, from healthcare robotics to autonomous transportation systems, becomes increasingly pronounced. However, continued vigilance and rigor in addressing the challenges associated with these technologies will be crucial in fully unleashing their potential and in paving the way for future research that pushes the boundaries of what embodied AI can accomplish.

7.2 Interdisciplinary Collaborations

The evolving field of Embodied Artificial Intelligence (Embodied AI) stands at the intersection of multiple scientific domains, each contributing unique insights and methodologies. This subsection explores the role of interdisciplinary collaborations as a catalyst for innovation in Embodied AI, demonstrating how merging expertise from various fields can address the intrinsic complexities of developing intelligent embodied systems.

Embodied AI inherently demands integration across diverse sciences, aiming to create intelligent systems capable of navigating and interacting with the real world in ways that closely approximate human capabilities. Central to this endeavor is the synthesis of insights from fields such as neuroscience, cognitive science, robotics, and computer science. Neuroscience provides a foundational understanding of the biological processes underlying human perception and action, inspiring the development of biologically plausible AI models [102]. For instance, principles of sensory integration and neural processing inform the design of sensorimotor learning algorithms that blend inputs from visual, auditory, and tactile senses, enhancing learning and interaction capabilities [103].

In parallel, cognitive sciences offer valuable perspectives on human cognition and behavior, essential for designing AI systems that exhibit attributes of learning, reasoning, and decision-making akin to humans. Cognitive architectures in AI often incorporate theories from cognitive science to model complex behaviors and adaptive learning processes [104]. As cognitive science continues to illuminate intelligence mechanisms, these insights are increasingly translated into computational models, driving the development of embodied agents capable of nuanced interactions within their environments.

Robotics, arguably the most closely allied field, provides the platform for implementing and testing Embodied AI systems. Innovations in soft robotics, control systems, and actuators facilitate the development of more dynamic and responsive robots, mimicking the agility and adaptability of biological organisms [105]. The application of reinforcement learning in robotics has enabled agents to learn from interactions and continuous feedback from their surroundings, incrementally refining their behaviors [106].

Moreover, computer science, particularly in machine learning and artificial intelligence, provides the computational backbone for these interdisciplinary endeavors. Techniques like multimodal machine learning integrate data across different sensory modalities, promoting a holistic perception framework crucial for embodied cognition [107]. This approach allows AI systems to process complex, heterogeneous data formats, paralleling the human ability to synthesize information from diverse sources.

Despite these advances, interdisciplinary collaborations face significant challenges. One major hurdle is integrating diverse methodologies and technologies into a cohesive framework. For example, aligning fast-paced developments in neuromorphic computing, which seeks to emulate neural architectures in hardware systems, with current AI models for efficient computation is a substantial challenge [88]. Additionally, ensuring the ethical alignment and societal acceptance of these technologies requires contributions from social sciences and ethical frameworks to guide responsible AI development.

Emerging trends in interdisciplinary collaborations point toward a more unified approach to AI development. Integrating virtual, augmented, and mixed reality technologies into Embodied AI research is opening new avenues for simulating and testing AI systems in controlled yet dynamic environments [87]. These innovations not only facilitate the development of more robust and adaptable AI agents but also ensure effective functioning in uncertain real-world conditions.

To further foster interdisciplinary collaborations, future research should consider establishing collaborative networks and platforms for data and knowledge exchange across disciplines. These networks could support the sharing of standardized datasets and simulation tools, accelerating innovation. Additionally, educational programs emphasizing interdisciplinary learning can prepare the next generation of researchers to address the complex challenges of Embodied AI.

In summary, interdisciplinary collaborations are key to unlocking the full potential of Embodied AI. By leveraging each field's strengths and addressing integration challenges, researchers can develop intelligent systems that are more capable and aligned with human values and societal needs. These collaborations, supported by robust frameworks and continuous dialogue between disciplines, promise to drive future innovations in Embodied AI, paving the way for systems that seamlessly integrate into human environments and enhance quality of life. The ongoing dialogue between theory and application will continue to illuminate the path for Embodied AI, progressing towards achieving its ultimate goal of Artificial General Intelligence (AGI).

7.3 Long-term Vision and Societal Impact

The long-term vision for embodied artificial intelligence (Embodied AI) is intricately linked with its transformative potential across technological and societal landscapes. As embodied systems integrate cognitive architectures with physical interactivity, they are poised to shift paradigms in human-technology coexistence, redefining interfaces in a myriad of domains ranging from healthcare to autonomous systems. This subsection aims to explore the anticipated trajectory of Embodied AI and its profound societal implications, offering a critical analysis of current approaches while identifying future challenges and opportunities.

At the forefront of Embodied AI's long-term development is the enhancement of trust and safety frameworks, ensuring that these systems not only understand but also adhere to societal norms and ethical standards while executing complex tasks. Trust in embodied systems is pivotal, especially as they become more autonomous and operated in sensitive areas such as caregiving and defense. This necessitates a robust framework for ensuring their reliability and ethical alignment. For example, the potential of integrating neuromorphic computing for real-time adaptation and reaction [24] offers a pathway towards systems better aligned with human cognition, driving towards safety and transparency in AI actions.

Further, the advancement of frameworks for open-ended task-solving capabilities in embodied systems provides a solid foundation for enhancing their adaptability and utility in real-world settings. This dynamic adaptability is crucial for addressing tasks without a clear structure or predefined outcomes, leveraging mechanisms such as reinforcement learning to inform decision-making processes [48]. Such capabilities underscore the transformative potential of embodied AI in tasks previously limited by rigid automation.

A visionary prospect for Embodied AI lies in fostering symbiotic human-machine collaboration, where embodied agents seamlessly integrate into human environments. This is predicated on the development of technologies that empower robots to not only perceive but also intuitively respond to human behaviors. For instance, the integration of emotional and empathic AI designs, as exemplified in works involving humanoid applications [91], heralds a future where machines are not merely tools but empathic collaborators. This evolution necessitates overcoming significant challenges, such as effective real-time processing and nuance understanding of human intent and environmental contexts.

A significant trend involves the co-optimization of control and morphology in soft embodied systems, offering iterations that mimic biological adaptations more closely. This entails the simultaneous evolution of both the physical and functional aspects of robots, facilitating more natural and adaptive interactions within varied environments [6]. However, a core challenge of this approach remains the complexity involved in aligning control systems with rapidly changing morphological states, demanding sophisticated evolutionary algorithms capable of balancing trade-offs between robustness and flexibility.

The societal impact of Embodied AI also extends to ethical and cultural dimensions. The deployment of these systems in everyday applications is poised to redefine social interactions, raising intricate questions about privacy, security, and equity. Recent advances in sensorimotor learning that enhance perceptual accuracy and environmental interaction [25] highlight the dual necessity of technical performance and ethical foresight. As we progress toward more human-like machines capable of autonomous decision-making, there is a pressing need for adaptable legal frameworks and consensus on the ethical boundaries of AI actions.

Moreover, interdisciplinary collaborations are pivotal, bridging insights from neuroscience, robotics, and social sciences to develop AI systems that not only perform tasks efficiently but do so in a manner that respects human values and societal norms. Studies have underscored the utility of embedding sensorimotor loops within embodied systems, enabling a nuanced understanding of environment dynamics [5]. This synthesis of disciplines contributes to a holistic approach to AI system design, fostering a more profound and socially aligned embodiment of intelligence.

In conclusion, the evolution of embodied AI hinges on the delicate balance of technological advancements with ethical and societal considerations. The anticipated trajectory will likely see these systems becoming tightly integrated into human life, necessitating stringent safety measures, adaptive frameworks for cohabitation, and robust interdisciplinary efforts. The potential for enhanced human-machine collaboration promises transformative impacts, yet the path forward must be navigated with care, ensuring that the societal benefits of such advancements are equitably realized. As embodied systems continue to evolve, it is imperative to maintain a forward-looking vision grounded in both technological prowess and ethical clarity, fostering an era where AI augments human potential while safeguarding societal values.

7.4 Evaluation and Benchmarking

In the domain of Embodied Artificial Intelligence (EAI), the evaluation and benchmarking of these systems are critical to ensuring robust and reliable performance in diverse real-world scenarios. This subsection explores the evolving methodologies designed to assess the performance, safety, and adaptability of EAI systems, building on the long-term vision of integrating embodied technologies into societal norms, as discussed in the previous sections. As EAI systems become increasingly entwined with daily functions, establishing comprehensive and standardized benchmarks is essential for fostering trust and advancing development in the field.

Evaluating EAI systems necessitates a multi-faceted approach due to the variety of capabilities these systems must exhibit. A significant trend is the shift toward comprehensive evaluation metrics that consider not only the functional performance of EAI systems but also their interaction capabilities, adaptability, safety, and ethical compliance. The integration of neuro-symbolic models, which blend neural perception with symbolic reasoning, represents a step towards more transparent and accountable AI practices by combining rigorous performance metrics across both neural and symbolic domains [108], [109].

Traditional benchmarks often lack the depth needed to capture the complexity of real-world environments in which EAI systems operate. Emerging efforts focus on simulating a range of conditions and scenarios, as exemplified by the Animal-AI Environment, which tests for animal-like cognition and behavioral adaptability, addressing complexities beyond the scope of simpler benchmarks [31]. Additionally, leveraging simulation environments with high fidelity in replicating real-world physics and procedural content generation aids in evaluating system effectiveness and resilience when transitioning from virtual to physical settings.

Bridging the sim-to-real gap remains a significant challenge, addressed by methodologies like domain randomization and online adaptation, which help models trained in simulations perform successfully in the physical world. These strategies involve varying training parameters in simulations to enhance generalization skills in unregulated environments. Evaluating these sim-to-real methodologies necessitates precise transfer evaluation metrics, assessing adaptability during deployment and sustained performance post-implementation.

An essential aspect of benchmarking EAI systems is employing standardized datasets that support fair evaluation and result reproducibility. The development of unified data standards allows for consistent assessments across various applications, facilitating comparative studies crucial for the field's progress. These datasets must cover a wide array of scenarios and variables reflective of the diverse nature of human environments, a challenge being addressed through collaborative research to establish universal evaluation standards.

Increasingly, the evaluation of social cognition and interactions in EAI systems is gaining attention, particularly within the context of human-robot interaction (HRI). The relational complexities inherent in social dynamics require specialized benchmarks to assess the social and emotional intelligence of EAI systems. Such benchmarks evaluate a system's ability to understand and respond to human emotions, social cues, and non-verbal communication, ensuring interactions remain natural and coherent [95]. Frameworks developed for these assessments often incorporate both quantitative and qualitative metrics, considering user satisfaction and acceptance, especially in sensitive applications like healthcare and caregiving [27].

Innovative approaches, such as the Dolores test, propose a Turing-like framework to evaluate whether embodied AI can seamlessly integrate emotional and cognitive faculties, demonstrating not only task efficiency but emotional fluency and adaptability as well [93]. Such frameworks highlight the importance of embedding emotional processing into AI core architecture, significantly enhancing human-machine collaboration and interaction.

In conclusion, the field of embodied AI evaluation and benchmarking is rapidly advancing, making substantial progress toward developing comprehensive benchmarks that address these systems' multi-dimensional capabilities. Future directions may include leveraging advancements in neuro-symbolic AI to further refine evaluation frameworks, ensuring that AI systems are not only functional and efficient but also relatable, understandable, and safe. Continuing interdisciplinary collaboration will be crucial to ensure that evaluation frameworks remain robust, comprehensive, and reflective of real-world complexities as part of the broader trajectory for Embodied AI's development and societal integration.

8 Conclusion

Embodied Artificial Intelligence (Embodied AI) represents a transformative frontier in the pursuit of creating systems capable of interacting with the world in human-like ways. As we conclude this comprehensive exploration of the field, it is imperative to synthesize the insights gained and chart out the future directions for research and application.

The convergence of core technologies, including sensory integration, motor control, and cognitive architectures, elucidates the potential of embodied systems to surpass traditional AI paradigms by grounding cognition in physical interactions [41]. Embodied AI's ability to incorporate environmental contexts into decision-making processes ensures enhanced situational awareness and adaptability compared to purely computational models [2]. This embodiment-driven approach reinforces the significance of morphological computation—where the physical form profoundly influences cognitive processes—thereby promoting efficient and effective behavior generation [1].

A comparative analysis of embodied systems reveals diverse methodologies, each with distinct strengths and limitations. For instance, paradigms such as Deep Evolutionary Reinforcement Learning and morphological innovation protection present innovative methods to co-optimize morphology and control, thus promoting adaptability in complex environments [6], [38]. In contrast, symbolic and sub-symbolic integration in cognitive architectures enables agents to traverse high-level reasoning and low-level sensorimotor tasks seamlessly [3]. Nonetheless, challenges persist, particularly in balancing computational efficiency with real-world scalability—a feat that remains central to advancing embodied learning and interaction strategies [2].

Furthermore, the transition from simulation to real-world application, known as the sim-to-real transfer, remains a critical milestone. While platforms like Habitat and systems such as RoboTHOR address simulation complexities, significant gaps still exist in ensuring the seamless deployment of simulation-trained models in physical settings [7], [10]. Continued innovation in domain randomization and physics-informed models is essential to bridge this chasm, ensuring embodied agents perform reliably in unpredictable environments.

As embodied AI continues its march towards Artificial General Intelligence (AGI), ethical, societal, and safety implications warrant acute consideration. Embodied conversational agents necessitate frameworks that ensure safe, transparent, and collaborative human-agent interactions [63], [110]. The integration of emotional and social cognition into embodied systems highlights the need for machines to contextualize and adapt to human non-verbal cues, thus fostering deeper trust and usability [62]. Addressing these ethical challenges, including bias mitigation and privacy enhancement measures, is vital for the responsible deployment of embodied AI technologies [111].

The future of embodied AI research is poised to benefit from interdisciplinary collaborations, drawing insights from neuroscience, psychology, and cognitive sciences to advance agent capabilities further [112]. Innovative frameworks such as the Technological Approach to Mind Everywhere (TAME) provide a blueprint for understanding diverse cognitive architectures, fostering more robust and flexible agents [20]. Moreover, emerging trends spotlight the potential of neuromorphic AI and spiking neural networks to offer bio-plausible solutions for efficient and scalable embodied intelligence [15].

In synthesis, the trajectory of Embodied AI necessitates relentless exploration in co-designing spatial AI systems, enhancing agent-environment interactions, and refining ethical, societal, and regulatory frameworks to align with rapid technological advancements [64]. As the landscape evolves, the integration of foundational models with embodied agents offers a promising avenue to bridge the gap between cyber and physical worlds, facilitating the development of robust AGI capable of complex real-world task execution.

In summation, the transformative potential of Embodied AI lies in its capacity to revolutionize machine cognition through the embodiment of physical, cognitive, and social capabilities. Continued exploration and collaboration across disciplines, coupled with an earnest commitment to ethical design and real-world applicability, will not only refine our scientific understanding but also imbue AI systems with human-like intelligence, autonomy, and empathy in increasingly dynamic environments.

References

[1] K. Zahedi and N. Ay, "Quantifying morphological computation," *ArXiv*, vol. abs/1301.6975, 2013. 1, 20, 25

[2] C. Moulin-Frier, J. Puigbò, X. Arsiwalla, M. Sánchez-Fibla, and P. Verschure, "Embodied artificial intelligence through distributed adaptive control: An integrated framework," *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 324–330, 2017. 1, 3, 6, 15, 19, 25

[3] M. Hoffmann and R. Pfeifer, "The implications of embodiment for behavior and cognition: animal and robotic case studies," *ArXiv*, vol. abs/1202.0440, 2012. 1, 3, 7, 10, 12, 16, 19, 21, 25

[4] T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh, "Symbol emergence in robotics: a survey," *Advanced Robotics*, vol. 30, pp. 706 – 728, 2015. 1, 11, 19

[5] G. Montúfar, K. Zahedi, and N. Ay, "A theory of cheap control in embodied systems," *PLoS Computational Biology*, vol. 11, 2014. 1, 3, 10, 12, 15, 24

[6] N. Cheney, J. Bongard, V. SunSpiral, and H. Lipson, "Scalable co-optimization of morphology and control in embodied machines," *Journal of The Royal Society Interface*, vol. 15, 2017. 1, 4, 14, 20, 22, 24, 25

[7] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, D. Parikh, and D. Batra, "Habitat: A platform for embodied ai research," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9338–9346, 2019. 2, 14, 22, 25

[8] Y. Liu, W. Chen, Y. Bai, J.-H. Luo, X. Song, K. Jiang, Z. Li, G. Zhao, J. Lin, G. Li, W. Gao, and L. Lin, "Aligning cyber space with physical world: A comprehensive survey on embodied ai," *ArXiv*, vol. abs/2407.06886, 2024. 2, 18

[9] N. Bredèche, E. Haasdijk, and A. Prieto, "Embodied evolution in collective robotics: A review," *Frontiers in Robotics and AI*, vol. 5, 2017. 2, 7

[10] M. Deitke, W. Han, A. Herrasti, A. Kembhavi, E. Kolve, R. Mottaghi, J. Salvador, D. Schwenk, E. VanderBilt, M. Wallingford, L. Weihs, M. Yatskar, and A. Farhadi, "Robothor: An open simulation-to-real embodied ai platform," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3161–3171, 2020. 2, 14, 22, 25

[11] J. Yang, Z. Ren, M. Xu, X. Chen, D. J. Crandall, D. Parikh, and D. Batra, "Embodied visual recognition," *ArXiv*, vol. abs/1904.04404, 2019. 2, 18

[12] C. Gan, Y. Gu, S. Zhou, J. Schwartz, S. Alter, J. Traer, D. Gutfreund, J. Tenenbaum, J. H. McDermott, and A. Torralba, "Finding fallen objects via asynchronous audio-visual integration," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10 513–10 523, 2022. 2

[13] P. D. Loor, K. Manac'h, and J. Tisseau, "Enaction-based artificial intelligence: Toward co-evolution with humans in the loop," *Minds and Machines*, vol. 19, pp. 319–343, 2009. 2, 6, 10, 11

[14] M. Ding, Y. Xu, Z. Chen, D. D. Cox, P. Luo, J. Tenenbaum, and C. Gan, "Embodied concept learner: Self-supervised learning of concepts and mapping through instruction following," *ArXiv*, vol. abs/2304.03767, 2023. 2

[15] R. V. W. Putra, A. Marchisio, F. Zayer, J. Dias, and M. Shafique, "Embodied neuromorphic artificial intelligence for robotics: Perspectives, challenges, and research development stack," *ArXiv*, vol. abs/2404.03325, 2024. 2, 18, 22, 26

[16] D. Kiela, L. Bulat, A. Vero, and S. Clark, "Virtual embodiment: A scalable long-term strategy for artificial intelligence research," *ArXiv*, vol. abs/1610.07432, 2016. 2, 17, 21

[17] P. Mazzaglia, T. Verbelen, B. Dhoedt, A. C. Courville, and S. Rajeswar, "Multimodal foundation world models for generalist embodied agents," *ArXiv*, vol. abs/2406.18043, 2024. 3

[18] Y. Wi, A. Zeng, P. R. Florence, and N. Fazeli, "Virdo++: Real-world, visuo-tactile dynamics and perception of deformable objects," in *Conference on Robot Learning*, 2022, pp. 1806–1816. 3

[19] G. Oliver, P. Lanillos, and G. Cheng, "Active inference body perception and action for humanoid robots," *ArXiv*, vol. abs/1906.03022, 2019. 3

[20] M. Levin, "Technological approach to mind everywhere: An experimentally-grounded framework for understanding diverse bodies and minds," *Frontiers in Systems Neuroscience*, vol. 16, 2021. 3, 26

[21] S. Kriegman, A. M. Nasab, D. S. Shah, H. Steele, G. Branin, M. Levin, J. Bongard, and R. Kramer-Bottiglio, "Scalable sim-to-real transfer of soft robot designs," *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 359–366, 2019. 4, 8, 12, 16

[22] J. Merel, A. Ahuja, V. Pham, S. Tunyasuvunakool, S. Liu, D. Tirumala, N. Heess, and G. Wayne, "Hierarchical visuomotor control of humanoids," *ArXiv*, vol. abs/1811.09656, 2018. 4, 8

[23] J. Merel, L. Hasenclever, A. Galashov, A. Ahuja, V. Pham, G. Wayne, Y. Teh, and N. Heess, "Neural probabilistic motor primitives for humanoid control," *ArXiv*, vol. abs/1811.11711, 2018. 4

[24] C. Richter, S. Jentzsch, R. Hostettler, J. Garrido, E. Vidal, A. Knoll, F. Röhrbein, P. van der Smagt, and J. Conradt, "Musculoskeletal robots: Scalability in neural control," *IEEE Robotics & Automation Magazine*, vol. 23, pp. 128–137, 2016. 4, 8, 16, 24

[25] K. Zahedi, D. Haeufle, G. Montúfar, S. Schmitt, and N. Ay, "Evaluating morphological computation in muscle and dc-motor driven models of hopping movements," *ArXiv*, vol. abs/1512.00250, 2015. 4, 12, 16, 24

[26] C. Moulin-Frier, T. Fischer, M. Petit, G. Pointeau, J. Puigbò, U. Pattacini, S. C. Low, D. Camilleri, P. D. H. Nguyen, M. Hoffmann, H. Chang, M. Zambelli, A.-L. Mealier, A. C. Damianou, G. Metta, T. Prescott, Y. Demiris, P. F. Dominey, and P. Verschure, "Dac-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, pp. 1005–1022, 2017. 4, 13

[27] H. Abdollahi, M. Mahoor, R. Zandie, J. Siewierski, and S. Qualls, "Artificial emotional intelligence in socially assistive robots for older adults: A pilot study," *IEEE Transactions on Affective Computing*, vol. 14, pp. 2020–2032, 2022. 4, 14, 21, 25

[28] P. Vossen, S. Baez, L. Bajcetic, and B. Kraaijeveld, "Leolani: a reference machine with a theory of mind for social communication," in *International Conference on Text, Speech and Dialogue*, 2018, pp. 15–25. 5

[29] A. N. Sheth, K. Roy, and M. Gaur, "Neurosymbolic ai - why, what, and how," *ArXiv*, vol. abs/2305.00813, 2023. 5, 9, 13, 17, 21

[30] A. Das, G. Gkioxari, S. Lee, D. Parikh, and D. Batra, "Neural modular control for embodied question answering," *ArXiv*, vol. abs/1810.11181, 2018. 5, 9, 13

[31] B. Beyret, J. Hernández-Orallo, L. Cheke, M. Halina, M. Shanahan, and M. Crosby, "The animal-ai environment: Training and testing animal-like artificial cognition," *ArXiv*, vol. abs/1909.07483, 2019. 5, 17, 25

[32] T. Wang, X. Mao, C. Zhu, R. Xu, R. Lyu, P. Li, X. Chen, W. Zhang, K. Chen, T. Xue, X. Liu, C. Lu, D. Lin, and J. Pang, "Embodiedscan: A holistic multi-modal 3d perception suite towards embodied ai," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19 757–19 767, 2023. 5, 15

[33] R. J. Savery, R. Rose, and G. Weinberg, "Establishing human-robot trust through music-driven robotic emotion prosody and gesture," *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–7, 2019. 5

[34] B. Jiang, X. Chen, C. Zhang, F. Yin, Z. Li, G. Yu, and J. Fan, "Motionchain: Conversational motion controllers via multimodal prompts," *ArXiv*, vol. abs/2404.01700, 2024. 5

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. 5

[36] Z. Cheng, Z.-Q. Cheng, J.-Y. He, J. Sun, K. Wang, Y. Lin, Z. Lian, X. Peng, and A. G. Hauptmann, "Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning," *ArXiv*, vol. abs/2406.11161, 2024. 5, 18, 22

[37] A. Szot, B. Mazoure, H. Agrawal, D. Hjelm, Z. Kira, and A. Toshev, "Grounding multimodal large language models in actions," *ArXiv*, vol. abs/2406.07904, 2024. 6

[38] A. Gupta, S. Savarese, S. Ganguli, and L. Fei-Fei, "Embodied intelligence via learning and evolution," *Nature Communications*, vol. 12, 2021. 6, 18, 23, 25

[39] J. M. Francis, N. Kitamura, F. Labelle, X. Lu, I. Navarro, and J. Oh, "Core challenges in embodied vision-language planning," *J. Artif. Intell. Res.*, vol. 74, pp. 459–515, 2021. 6

[40] Y. Ma, Z. Song, Y. Zhuang, J. Hao, and I. King, "A survey on vision-language-action models for embodied ai," *ArXiv*, vol. abs/2405.14093, 2024. 6, 18

[41] J. Duan, S. Yu, T. Li, H. Zhu, and C. Tan, "A survey of embodied ai: From simulators to research tasks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, pp. 230–244, 2021. 6, 18, 25

[42] Y. Long, W. Wei, T. Huang, Y. Wang, and Q. Dou, "Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning," *IEEE Robotics and Automation Letters*, vol. 8, pp. 4441–4448, 2023. 6, 14

[43] M. Savva, A. X. Chang, A. Dosovitskiy, T. Funkhouser, and V. Koltun, "Minos: Multimodal indoor simulator for navigation in complex environments," *ArXiv*, vol. abs/1712.03931, 2017. 7, 15

[44] C. Chen, U. Jain, C. Schissler, S. V. A. Garí, Z. Al-Halah, V. Ithapu, P. Robinson, and K. Grauman, "Soundspaces: Audio-visual navigation in 3d environments," in *European Conference on Computer Vision*, 2019, pp. 17–36. 7, 15

[45] S. Heinrich and S. Wermter, "Interactive natural language acquisition in a multi-modal recurrent neural architecture," *Connection Science*, vol. 30, pp. 133 – 99, 2017. 7, 19

[46] S. Thermos, G. Papadopoulos, P. Daras, and G. Potamianos, "Deep affordance-grounded sensorimotor object recognition," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 49–57, 2017. 7, 11, 15

[47] L. Chen, Y. Zhang, S. Ren, H. Zhao, Z. Cai, Y. Wang, T. Liu, and B. Chang, "Towards end-to-end embodied decision making via multi-modal large language model: Explorations with gpt4-vision and beyond," *ArXiv*, vol. abs/2310.02071, 2023. 7

[48] A. Dosovitskiy and V. Koltun, "Learning to act by predicting the future," *ArXiv*, vol. abs/1611.01779, 2016. 8, 12, 24

[49] Y. Tassa, S. Tunyasuvunakool, A. Muldal, Y. Doron, S. Liu, S. Bohez, J. Merel, T. Erez, T. Lillicrap, and N. Heess, "dm_control: Software and tasks for continuous control," *Softw. Impacts*, vol. 6, p. 100022, 2020. 8, 16, 20

[50] T. Howison, S. Hauser, J. Hughes, and F. Iida, "Reality-assisted evolution of soft robots through large-scale physical experimentation: A review," *Artificial Life*, vol. 26, pp. 484–506, 2020. 8

[51] Y.-L. Qiao, J. Liang, V. Koltun, and M. C. Lin, "Efficient differentiable simulation of articulated bodies," *ArXiv*, vol. abs/2109.07719, 2021. 8

[52] M. Hersche, F. di Stefano, T. Hofmann, A. Sebastian, and A. Rahimi, "Probabilistic abduction for visual abstract reasoning via learning rules in vector-symbolic architectures," *ArXiv*, vol. abs/2401.16024, 2024. 9

[53] W. Wang, Y. Yang, and F. Wu, "Towards data-and knowledge-driven artificial intelligence: A survey on neuro-symbolic computing," 2022. 9

[54] A. Bera, T. Randhavane, R. Prinja, K. Kapsaskis, A. Wang, K. Gray, and D. Manocha, "The emotionally intelligent robot: Improving social navigation in crowded environments," *ArXiv*, vol. abs/1903.03217, 2019. 9, 13

[55] S. Parekh and D. P. Losey, "Learning latent representations to co-adapt to humans," *Autonomous Robots*, vol. 47, pp. 771 – 796, 2022. 9, 13

[56] P. Fung, D. Bertero, Y. Wan, A. Dey, R. Chan, F. B. Siddique, Y. Yang, C.-S. Wu, and R. Lin, "Towards empathetic human-robot interactions," in *Conference on Intelligent Text Processing and Computational Linguistics*, 2016, pp. 173–193. 9, 13, 21

[57] A. Ghandeharioun, D. J. McDuff, M. Czerwinski, and K. Rowan, "Emma: An emotion-aware wellbeing chatbot," *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, 2018. 9, 10

[58] P. Jonell, T. Kucherenko, G. Henter, and J. Beskow, "Let's face it: Probabilistic multi-modal interlocutor-aware generation of facial gestures in dyadic settings," *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 2020. 9

[59] Y. Hong, Z. Zheng, P. Chen, Y. Wang, J. Li, and C. Gan, "Multiply: A multisensory object-centric embodied large language model in 3d world," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26 396–26 406, 2024. 10

[60] A. Netanyahu, T. Shu, B. Katz, A. Barbu, and J. Tenenbaum, "Phase: Physically-grounded abstract social events for machine social perception," *ArXiv*, vol. abs/2103.01933, 2021. 10, 14, 18

[61] P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions," *ArXiv*, vol. abs/2209.03430, 2022. 10, 17, 18

[62] J. Bütepage and D. Kragic, "Human-robot collaboration: From psychology to social robotics," *ArXiv*, vol. abs/1705.10146, 2017. 10, 14, 25

[63] E. Deng, B. Mutlu, and M. Matarić, "Embodiment in socially interactive robots," *Found. Trends Robotics*, vol. 7, pp. 251–356, 2019. 10, 21, 25

[64] D. Bohus, S. Andrist, A. Feniello, N. Saw, M. Jalobeanu, P. Sweeney, A. L. Thompson, and E. Horvitz, "Platform for situated intelligence," *ArXiv*, vol. abs/2103.15975, 2021. 10, 18, 26

[65] C. Zhang, Z. Yang, X. He, and L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, pp. 478–493, 2019. 11

[66] J. Dávila-Chacón, J. Liu, and S. Wermter, "Enhanced robot speech recognition using biomimetic binaural sound source localization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 138–150, 2019. 11, 19

[67] C. Matuszek, N. FitzGerald, L. Zettlemoyer, L. Bo, and D. Fox, "A joint model of language and perception for grounded attribute learning," in *International Conference on Machine Learning*, 2012, pp. 1435–1442. 11

[68] N. Heess, G. Wayne, Y. Tassa, T. Lillicrap, M. A. Riedmiller, and D. Silver, "Learning and transfer of modulated locomotor controllers," *ArXiv*, vol. abs/1610.05182, 2016. 12, 16

[69] D. Arumugam, S. Karamcheti, N. Gopalan, L. L. S. Wong, and S. Tellex, "Accurately and efficiently interpreting human-robot instructions of varying granularities," *ArXiv*, vol. abs/1704.06616, 2017. 13

[70] M. M. Scheunemann, R. Cuijpers, and C. Salge, "Warmth and competence to predict human preference of robot behavior in physical human-robot interaction," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1340–1347, 2020. 13

[71] M. Spitale and H. Gunes, "Affective robotics for wellbeing: A scoping review," *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1–8, 2022. 13

[72] L. Zhang, Z. Ji, and B. Chen, "Crew: Facilitating human-ai teaming research," *ArXiv*, vol. abs/2408.00170, 2024. 13

[73] K. Acharya, W. Raza, C. Dourado, A. Velasquez, and H. Song, "Neurosymbolic reinforcement learning and planning: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 5, pp. 1939–1953, 2023. 13, 21

[74] I. Gaudiello, E. Zibetti, S. Lefort, M. Chetouani, and S. Ivaldi, "Trust as indicator of robot functional and social acceptance. an experimental study on user conformation to icub answers," *Comput. Hum. Behav.*, vol. 61, pp. 633–655, 2015. 13, 17, 20

[75] P. Wolfert, N. L. Robinson, and T. Belpaeme, "A review of evaluation practices of gesture generation in embodied conversational agents," *IEEE Transactions on Human-Machine Systems*, vol. 52, pp. 379–389, 2021. 13

[76] B. L. Bhatnagar, X. Xie, I. A. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll, "Behave: Dataset and method for tracking human object interactions," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15 914–15 925, 2022. 14

[77] J. Abou-Chakra, K. Rana, F. Dayoub, and N. Sunderhauf, "Physically embodied gaussian splatting: A realtime correctable world model for robotics," *ArXiv*, vol. abs/2406.10788, 2024. 15

[78] C. Gan, J. Schwartz, S. Alter, M. Schrimpf, J. Traer, J. D. Freitas, J. Kubilius, A. Bhandwaldar, N. Haber, M. Sano, K. Kim, E. Wang, D. Mrowca, M. Lingelbach, A. Curtis, K. T. Feigelis, D. Bear, D. Gutfreund, D. Cox, J. DiCarlo, J. H. McDermott, J. Tenenbaum, and D. L. K. Yamins, "Threedworld: A platform for interactive multi-modal physical simulation," *ArXiv*, vol. abs/2007.04954, 2020. 15

[79] Y. Yuan and K. Kitani, "Residual force control for agile human behavior imitation and extended motion synthesis," *ArXiv*, vol. abs/2006.07364, 2020. 16, 20

[80] C. B. Schaff, A. Sedal, and M. R. Walter, "Soft robots learn to crawl: Jointly optimizing design and control with sim-to-real transfer," *ArXiv*, vol. abs/2202.04575, 2022. 16

[81] C. Zhang, W. Xiao, T. He, and G. Shi, "Wococo: Learning whole-body humanoid control with sequential contacts," *ArXiv*, vol. abs/2406.06005, 2024. 16

[82] D. Pathak, C. Lu, T. Darrell, P. Isola, and A. A. Efros, "Learning to control self-assembling morphologies: A study of generalization via modularity," *ArXiv*, vol. abs/1902.05546, 2019. 16

[83] V. Narayanan, B. M. Manoghar, V. S. Dorbala, D. Manocha, and A. Bera, "Proxemo: Gait-based emotion learning and multi-view proxemic fusion for socially-aware robot navigation," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8200–8207, 2020. 17

[84] C. Brooks and D. Szafir, "Building second-order mental models for human-robot interaction," *ArXiv*, vol. abs/1909.06508, 2019. 17

[85] J. Chen, C. Gao, E. Meng, Q. Zhang, and S. Liu, "Reinforced structured state-evolution for vision-language navigation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15 429–15 438, 2022. 18

[86] L. Weihs, J. Salvador, K. Kotar, U. Jain, K.-H. Zeng, R. Mottaghi, and A. Kembhavi, "Allenact: A framework for embodied ai research," *ArXiv*, vol. abs/2008.12760, 2020. 19

[87] T. Groechel, M. Walker, C. T. Chang, E. Rosen, and J. Forde, "A tool for organizing key characteristics of virtual, augmented, and mixed reality for human–robot interaction systems: Synthesizing vam-hri trends and takeaways," *IEEE Robotics & Automation Magazine*, vol. 29, pp. 35–44, 2021. 19, 23

[88] G. Zardini, D. Milojevic, A. Censi, and E. Frazzoli, "Co-design of embodied intelligence: A structured approach," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7536–7543, 2020. 19, 23

[89] K. Kawaharazuka, K. Tsuzuki, S. Makino, M. Onitsuka, Y. Asano, K. Okada, K. Kawasaki, and M. Inaba, "Long-time self-body image acquisition and its application to the control of musculoskeletal structures," *IEEE Robotics and Automation Letters*, vol. 4, pp. 2965–2972, 2019. 19

[90] C. Lenz and S. Behnke, "Bimanual telemanipulation with force and haptic feedback through an anthropomorphic avatar system," *Robotics Auton. Syst.*, vol. 161, p. 104338, 2022. 20

[91] M. K. Mittal, D. Hoeller, F. Farshidian, M. Hutter, and A. Garg, "Articulated object interaction in unknown scenes with whole-body mobile manipulation," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1647–1654, 2021. 20, 24

[92] A. Cherpillod, S. Mintchev, and D. Floreano, "Embodied flight with a drone," *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pp. 386–390, 2017. 20

[93] L. Pessoa, "Intelligent architectures for robotics: The merging of cognition and emotion," *Physics of life reviews*, 2019. 20, 25

[94] W. Xu and Z. Gao, "Applying hcai in developing effective human-ai teaming: A perspective from human-ai joint cognitive systems," *Interactions*, vol. 31, pp. 32 – 37, 2023. 20

[95] R. J. Savery and G. Weinberg, "A survey of robotics and emotion: Classifications and models of emotional interaction," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 986–993, 2020. 21, 25

[96] T. Shu, A. Bhandwaldar, C. Gan, K. A. Smith, S. Liu, D. Gutfreund, E. Spelke, J. Tenenbaum, and T. Ullman, "Agent: A benchmark for core psychological reasoning," in *International Conference on Machine Learning*, 2021, pp. 9614–9625. 21

[97] I. Habibie, W. Xu, D. Mehta, L. Liu, H. Seidel, G. Pons-Moll, M. A. Elgharib, and C. Theobalt, "Learning speech-driven 3d conversational gestures from video," *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, 2021. 21

[98] H. Joo, T. Simon, M. Cikara, and Y. Sheikh, "Towards social artificial intelligence: Nonverbal social signal prediction in a triadic interaction," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10 865–10 875, 2019. 21

[99] A. S. Sundar and L. Heck, "Multimodal conversational ai: A survey of datasets and approaches," *ArXiv*, vol. abs/2205.06907, 2022. 22

[100] J. Abramson, A. Ahuja, F. Carnevale, P. Georgiev, A. Goldin, A. Hung, J. Landon, J. Lhotka, T. Lillicrap, A. Muldal, G. Powell, A. Santoro, G. Scully, S. Srivastava, T. von Glehn, G. Wayne, N. Wong, C. Yan, and R. Zhu, "Improving multimodal interactive agents with reinforcement learning from human feedback," *ArXiv*, vol. abs/2211.11602, 2022. 22

[101] D. Driess, F. Xia, M. S. M. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, W. Huang, Y. Chebotar, P. Sermanet, D. Duckworth, S. Levine, V. Vanhoucke, K. Hausman, M. Toussaint, K. Greff, A. Zeng, I. Mordatch, and P. R. Florence, "Palm-e: An embodied multimodal language model," in *International Conference on Machine Learning*, 2023, pp. 8469–8488. 22

[102] P. Lanillos and M. Gerven, "Neuroscience-inspired perception-action in robotics: applying active inference for state estimation, control and self-perception," *ArXiv*, vol. abs/2105.04261, 2021. 23

[103] S. Heinrich, Y. Yao, T. Hinz, Z. Liu, T. Hummel, M. Kerzel, C. Weber, and S. Wermter, "Crossmodal language grounding in an embodied neurocognitive model," *Frontiers in Neurorobotics*, vol. 14, 2020. 23

[104] F. Peller-Konrad, R. Kartmann, C. R. G. Dreher, A. Meixner, F. Reister, M. Grotz, and T. Asfour, "A memory system of a robot cognitive architecture and its implementation in armarx," *Robotics Auton. Syst.*, vol. 164, p. 104415, 2022. 23

[105] T. G. Thuruthel and F. Iida, "Multimodel sensor fusion for learning rich models for interacting soft robots," *ArXiv*, vol. abs/2205.04202, 2022. 23

[106] D. Palenicek, T. Gruner, T. Schneider, A. Böhm, J. Lenz, I. Pfenning, E. Krämer, and J. Peters, "Learning tactile insertion in the real world," *ArXiv*, vol. abs/2405.00383, 2024. 23

[107] I. Momennejad, "A rubric for human-like agents and neuroai," *Philosophical Transactions of the Royal Society B*, vol. 378, 2022. 23

[108] A. Garcez and L. Lamb, "Neurosymbolic ai: the 3rd wave," *Artificial Intelligence Review*, pp. 1–20, 2020. 24

[109] Z. Wan, C.-K. Liu, H. Yang, C. Li, H. You, Y. Fu, C. Wan, T. Krishna, Y. Lin, and A. Raychowdhury, "Towards cognitive ai systems: a survey and prospective on neuro-symbolic ai," *ArXiv*, vol. abs/2401.01040, 2024. 24

[110] G. Paolo, J. Gonzalez-Billandon, and B. K'egl, "A call for embodied ai," *ArXiv*, vol. abs/2402.03824, 2024. 25

[111] N. Roy, I. Posner, T. Barfoot, P. Beaudoin, Y. Bengio, J. Bohg, O. Brock, I. Depatie, D. Fox, D. Koditschek, T. Lozano-Perez, V. K. Mansinghka, C. Pal, B. Richards, D. Sadigh, S. Schaal, G. Sukhatme, D. Thérien, M. Toussaint, and M. V. D. Panne, "From machine learning to robotics: Challenges and opportunities for embodied intelligence," *ArXiv*, vol. abs/2110.15245, 2021. 25

[112] L. Zhao, L. Zhang, Z. Wu, Y. Chen, H. Dai, X.-X. Yu, Z. Liu, T. Zhang, X. Hu, X. Jiang, X. Li, D. Zhu, D. Shen, and T. Liu, "When brain-inspired ai meets agi," *ArXiv*, vol. abs/2303.15935, 2023. 25