Robotic Foundation Models Should Evolve Toward an Interactive Multi-Agent Perspective

Sharmita Dey*
ETH Zurich,
University of Goettingen
contact.deysharmita@gmail.com

Strahinja Dosen
Aalborg University

Stefano Albrecht
DeepFlow London

Abstract

Recent advances in large-scale machine learning have produced high-capacity *foundation models* capable of adapting to a wide range of downstream tasks. While such models hold great promise for robotics, the prevailing paradigm still portrays robots as single, autonomous decision-makers, performing tasks such as manipulation and navigation, with limited human involvement. However, a large class of real-world robotic systems, including wearable robotics (e.g., prostheses, orthoses, exoskeletons), teleoperation, and neural interfaces, are semiautonomous, and require ongoing interactive coordination with human partners, challenging single-agent assumptions. In this position paper, we argue that robot foundation models must evolve to an *interactive multi-agent* perspective in order to handle the complexities of real-time human-robot co-adaptation. To ground our discussion, we identify generalizable neuroscience-inspired functionalities required in such a multi-agent approach: (1) a multimodal *sensing module* informed by sensorimotor integration principles for collaborative sensing, (2) a teamwork model reminiscent of joint-action frameworks in cognitive science for collaborative actions, (3) a predictive world belief model grounded in internal forward model theories of motor control for anticipation and planning, and (4) a memory/feedback mechanism that echoes concepts of Hebbian and reinforcement-based plasticity for model refinement. By moving beyond single-agent perspective, our position emphasizes how foundation models in robotics can engage in adaptive interactions with humans and other agents, thereby enhancing their functionality and applicability in complex, dynamic environments.

1 Introduction

In recent years, artificial intelligence has been transformed by *foundation models*, which are large, high-capacity neural networks pre-trained on extensive and heterogeneous datasets [14, 2]. These models, exemplified by large language models (LLMs) such as GPT-4 [2], and large multimodal models (LMMs) like PaLM-E [45], offer a flexible interface for perception, reasoning, and action. In robotics, foundation models have been applied to unify diverse tasks, e.g., manipulation, navigation, or object recognition, under a single policy [103, 13], often operating in a *single-agent* paradigm where the robot acts autonomously, under minimal human involvement.

*corresponding author. SDey conceived the ideas and wrote the first draft. SAlbrecht and SDosen contributed domain expertise, literature; refined and revised the manuscript.

Preprint.

However, a large class of real-world robotics, particularly those involving continuous human collaboration, are inherently *multi-agent*. Applications such as teleoperation [76, 94, 68] and rehabilitation robotics, which includes prosthetic devices [10, 24, 99, 129, 42, 40] and exoskeletons [85, 104, 79, 31, 38, 29], neural interfaces [61, 44, 110, 126], brain-computer interfaces [130, 91, 80, 1], and other semi-autonomous systems [117, 25] require ongoing co-adaptation with humans or other participating agents in the environment, rather than isolated, one-shot instructions. In these contexts, the single-agent perspective encounters significant limitations: it fails to interpret and handle *non-stationary* factors such as dynamic and evolving user states, shifting goals, user fatigue, and changing environmental conditions.

Human-interactive robotics and bionics, in particular, demand continuous, bidirectional feedback loops between the human user and the device [67, 35]. They require a high degree of integration and coordination between the human and the robot. To achieve this, the device must anticipate human actions, integrate user preferences and environmental cues to ensure effective, safe, and comfortable completion of user commands. Over time, both the human and the device need to learn to function as a coordinated pair. Consequently, this paper takes a position: **future robot foundation models must evolve to an interactive multi-agent perspective**, by explicitly modeling both the robot and its counterpart (human or environment) as actively adapting agents. This complexity aligns more closely with *neuroscience-based* perspectives on sensorimotor control, which emphasize dynamic feedback loops, internal predictive models, and adaptive synergy between multiple interacting systems (e.g., brain, muscles, external supports) [133, 64]. By translating these concepts into modular components of a robotic system, we outline how an interactive human-robot dyad can be realized in future foundation models.

We illustrate this perspective mostly through wearable robotics or human bionic systems, where a user's actions and physiological signals (e.g., EMG, joint angles) need to continuously and seamlessly intertwine with the device's actuation. In general, the principles outlined here are broadly applicable to robots operating in semi-autonomous or interactive contexts. By incorporating principles from neuroscience, cognitive science, and multi-agent systems, this position paper aims to generate discussion on how to achieve co-adaptation, comfort, and anticipatory control in next-generation interactive robotics. We suggest that the widespread adoption of interactive, multi-agent paradigms in robotic foundation models will lead to fundamentally safer, more robust, and user-centric performance, surpassing what is possible within the single-agent paradigm.

2 Single-Agent Foundation Models in Robotics

The advent of Large Language Models (LLMs) such as GPT-4 [2], LLaMA [122], and Vision-Language Models (VLMs) such as CLIP [100], BLIP [74], BLIP-2 [73] has significantly advanced robotics by enhancing perception, planning, and action generation capabilities. These models demonstrate exceptional abilities in understanding and generating multimodal data, which are crucial for complex robotic tasks [45, 13, 120]. By leveraging the robust linguistic capabilities of LLMs, robots can interpret and execute tasks based on natural language commands [54], eliminating the need for complex programming interfaces. For instance, robots can parse instructions such as "Bring the red cup from the kitchen table" into structured subtasks involving object identification, navigation, and manipulation [14].

Despite the impressive capabilities of modern robot foundation models, such as Gato [103], RT-1 [13], RT-2 [12], RT-X [96], Octo [120], and OpenVLA [66] their predominantly single-agent framework can limit performance in scenarios requiring tight coordination with humans or other agents. These models typically learn policies under the assumption that the robot operates largely on its own, taking in sensory inputs and issuing motor commands without ongoing, interactive feedback from a collaborator or user. Although they have achieved notable results on tasks like manipulation, navigation, and even some language grounding, key shortcomings emerge when real-time and collaboration with the human operator or continuous human guidance is essential.

2.1 Limitations of Single-Agent Robot Foundation Models

1) Inability to Handle Mid-Task User Corrections and Human-Robot "Turn-Taking". Contemporary robot foundation policies (e.g. Gato, RT-1, RT-2, Octo) are trained to execute a given goal end-to-end without intermediate interaction. Gato [103], for example, was demonstrated in

multiple discrete and continuous control tasks, ranging from Atari gameplay to real-world robotic arm manipulation. However, it was not designed to handle situations where a human might intervene mid-task with corrective feedback or dynamically changing instructions (e.g., "Wait, do not place the block there, hand it to me instead."). As a single-agent learner, Gato or Octo follow their end-to-end policy after receiving an initial goal or observation. If the human's intent shifts *during* task execution, they cannot seamlessly incorporate that feedback without externally resetting or retraining the policy. A *multi-agent* perspective, by contrast, would treat the user as a parallel decision-maker; the system would maintain a belief state about the user's evolving instructions, thus adapting plans in real time rather than requiring full restarts.

2) Missed Opportunities for Co-Adaptation. Both Gato and RT-1 illustrate the single-agent assumption that the robot alone adapts its policy. In scenarios like teleoperation or assistive tasks, however, adaptation is a two-way street: the human also modifies their behavior in response to how the robot is acting, and vice versa. A single-agent viewpoint cannot fully leverage user posture shifts, subtle gestural cues, or real-time user feedback on comfort and safety. By contrast, a multi-agent approach (e.g., ad-hoc teamwork [101, 83]) would explicitly model how the human's internal state may change, whether due to fatigue, changing preferences, or partial completion of sub-goals and modify the robot's behavior accordingly. This co-adaptive loop can prevent errors (e.g., collisions, user frustration) that arise when the robot rigidly executes a policy absent mutual feedback.

3) Overlooking Collaborative Goal Setting and Preference Tracking. Another limitation is that single-agent foundation models rarely incorporate long-term *preference tracking* for an external user. For instance, neither Gato, RT-1, RT-2, nor Octo record that a particular user "prefers gentler grasps" or has a habit of signaling differently; every instruction is treated in isolation. While they excel at learning general policies from large datasets, they do not maintain a persistent model of a user's personal constraints or historical preferences (e.g., "User typically prefers lighter grip force on fragile objects" or "User signals discomfort when the end-effector approaches from the left"). In a multi-agent framework, the robot could treat the user's preferences as a dynamic factor, continually updating its internal representation as tasks progress and new user feedback arises, thus improving safety, utility and user satisfaction [72].

3 From Autonomous Single Agents to Multi-Agent Collaboration

To address limitations of single-agent paradigms, especially in environments that demand complex, dynamic interactions and collaborative problem solving, research has advanced toward *multi-agent systems* (MAS) [113], multi-agent reinforcement learning [3], and *ad-hoc teamwork* [83], frameworks designed to facilitate effective collaboration among multiple entities. Multi-agent systems research has long studied how agents can cooperate or compete without central coordination [113, 135, 3]. These agents can be homogeneous or heterogeneous, cooperative or competitive, and operate within shared or overlapping environments. The primary distinction between MAS and single-agent systems lies in the ability to manage interdependencies and leverage collective intelligence to solve problems that are intractable for individual agents. In particular, ad hoc teamwork research formalizes the challenge of rapidly adapting to unknown teammates and tasks without prior coordination [83, 116]. This capability is crucial in environments where agents must form temporary coalitions spontaneously to achieve common objectives, often under conditions of uncertainty and incomplete information [7, 81, 101].

Recently, multi-agent LLM frameworks have begun exploring how multiple large language models or "agents" can interact to solve tasks. For example, Microsoft's AutoGen framework composes multiple LLM-powered agents that converse with each other to accomplish goals [136]. AutoGen agents can be specialized (one handling math, another code, etc.) and coordinate via natural language. Similarly, the AutoAgents system [20] automatically generates a team of agents from a task description, each with different roles [123]. These and related platforms demonstrate how LLMs can be orchestrated into collaborative multi-agent pipelines. Indeed, recent surveys highlight that LLM-based MAS enable groups of agents to collectively perceive, reason, and act on complex problems [136, 123]. Beyond specific frameworks, many open-source libraries (e.g., LangChain agent chains [19], OpenAI's "Swarm" [95], etc.) and commercial tools support building multi-agent workflows where agents specialize or iteratively refine solutions. This trend suggests that intelligence is being distributed across LLM agents in an ad hoc fashion. However, most of this work has focused on textual or

planning tasks, not directly on embodied robotics. In particular, control of wearable robotics, a prime example of human-robot coupling, was mostly implemented by aiming to "reconnect" the robot to the sensorimotor structures of the human user. In this framework, the robot controller could be regarded as a "simple" decoder aiming to estimate user motion intention and translate it into robot commands [62]. More recently, smart bionic systems have emerged in which the controller possesses rudimentary reasoning abilities, allowing it to carry out certain tasks autonomously [50]. Thus far, however, these autonomous controllers have been realized solely as single-agent models.

3.1 Human-Robot Dyad as a Real-World Multi-Agent Challenge

A defining feature of human-interactive robots or bionic systems (i.e., wearable robotics, prostheses, and exoskeletons) is the *fusion* of human physiology with artificial actuation [8, 115, 124, 47, 34, 30, 32, 37, 36, 41, 39]. In these scenarios, the user (with biological muscles, joints, and neural control) and the robotic device (with actuators, sensors, and algorithms) act as two tightly coupled agents. As explained in prior sections, most existing approaches still adopt a single-agent viewpoint: the device waits for explicit commands and reacts in a largely feed-forward manner. This perspective ignores the subtle, bidirectional continual interplay that actually unfolds, an omission that can cause misinterpreted user inputs, abrupt control actions, delayed task transitions, or poorly personalised assistance when a user's biomechanical or cognitive state shifts unexpectedly [89].

Reframing the human and the device as two partially observable, co-evolving agents opens the door to modern multi-agent collaboration paradigms, wherein: 1) the device continuously estimates the user's goals and biomechanical limits from multimodal signals (e.g., EMG, limb kinematics). 2) The human updates their motor strategy in response to the device's feedback and behaviour, closing the loop in real time. 3) The device maintains and updates an internal model of the user's state and preferences, enabling more synergistic control over repeated interactions.

Even in ostensibly 'autonomous' robot applications [33, 43, 109, 86], there can be hidden interaction partners, such as a human operator providing commands or an environment whose states change in response to robot actions. Integrating multi-agent interactions into foundation models equips robots with explicit representations of user states, fostering the ability to *predict* and *adapt* continually based on teammates' models of the world. This integration addresses the fundamental limitations of single-agent, autonomous controllers in human-interactive settings.

4 Position: Embracing an Interactive Multi-Agent Foundational Architecture Inspired by Neuroscience

We argue that **future robot foundation models must adopt an interactive multi-agent framework**, especially for human-robot-interactive domains, one that recognizes the user and the robot (e.g., a smart robotic prosthetic device) as two interacting agents. To ground our discussion in concrete building blocks, we outline four desirable functional capabilities: (1) A *collaborative sensing module* that fuses *heterogeneous modalities* and *viewpoints*, such as, robot exteroception (camera, lidar, force), egocentric human signals (EMG, inertial units, eye-gaze), and sparse but semantically-rich cues (speech, gestures) into a compact, task-relevant latent representation, mirroring parietal sensorimotor integration in biological systems [98]. (2) A *teamwork modeling* module [101, 72, 83] that applies multi-agent collaboration principles, aligning with joint-action and shared intentionality theories in cognitive science [121, 111]. (3) A *predictive world belief model* [52, 71, 29, 35] that maintains an internal forward model of the user and/or collaborating agents' states, to simulate how the collaborative dyad and the environment are likely to evolve in the future, enabling anticipatory control, inspired by motor control theories on forward internal models and predictive coding [133, 132, 134, 102, 82, 28]. (4) A *memory/feedback mechanism* that stores user-specific preferences and updates policies in a reinforcement-like manner, similar to the role of synaptic plasticity and reinforcement learning in shaping long-term sensorimotor adaptations [27]. The high-level concept of integration of world-modeling and memory mechanisms has also been supported in broader cognitive architectures (e.g., [71]); however, in our work, this integration is situated within collaborative contexts.

Importantly, these are not a rigid pipeline; rather, we identify the core capabilities that a human-interactive system should possess. The main purpose of identifying these elements is to invite the attention of the researchers to the open research questions that need to be addressed for building future

foundation models for human-robot co-adaptation. For each of the elements, we draw neuroscientific parallels that can eventually guide in manifesting similar capabilities in artificial neural networks. Where applicable, we also discuss how the current literature in deep learning and robotics approach these questions and the gaps that need to be filled to integrate such solutions for human-robot collaboration. Thus, we aim to formalize the requirements of future robotic foundation models for human-robot collaboration and propose directions to achieve them.

4.1 Module 1: Multi-agent Collaborative Sensing

Sensing in a multi-agent environment begins with the premise that each agent, human or robotic, carries a partial, modality-specific slice of reality. The sensing module, therefore, must fuse both heterogeneous modalities and heterogeneous viewpoints into one temporally coherent belief that downstream teamwork, prediction, and memory mechanisms can trust. State-of-the-art cooperative perception offers many alternatives to achieve this in a multi-agent system. Reconstruction-driven methods [127] learn a shared latent that can rebuild the scene that each agent is missing. Sensor-fusion-based approaches formalize the information exchange in different collaboration stages [139, 59]: early fusion, where agents broadcast raw sensor streams; late fusion, where they exchange only high-level detections; intermediate fusion, where they trade deep-feature maps. Bandwidth-centric solutions such as When2Com [78], Where2Comm [57], PragComm [58], treat perception as a constrained communication problem and compress the shared information by learning to select what information to send and where without compromising accuracy.

Translating these ideas to human-robot interaction (HRI) requires a shift of perspective: the other agent's sensorium is biological, bandwidth-limited, and partly private. One approach to bridge this gap is sensor symmetry: equip the human collaborator with lightweight egocentric cameras [22], electro-physiological or motion sensors (e.g., EMG sleeves for a prosthesis user, inertial tags for a teleoperator, eye-trackers for shared mapping) so that their viewpoint can be time-stamped, spatially calibrated, and fused like any sensor stream [90]. A second approach is asymmetric fusion: treat language, gaze direction, or hand gestures as sparse but high-semantic tokens [128]; robots learn a task-aware dictionary that gives a few bits of human intent the same weight as hundreds of lidar voxels. Finally, implicit embedding approaches can infer human latent belief (e.g., intended obstacle positions) from motor commands and speech, folding that belief into shared predictive latent [56].

Neuroscience offers the organizing principle to achieve this: parietal circuits fuse visual, vestibular, and proprioceptive cues to reconcile egocentric and allocentric frames, forwarding only prediction, error signals to premotor and motor cortices for action selection [60, 119, 21]. An HRI sensing module should first align robot exteroception with human-embodied signals inside a multisensory hub, a common, uncertainty-aware body-centred schema, and then broadcast a light-weight latent vector, weighted by the task value of each cue, to the teamwork and prediction modules, mirroring sensorimotor integration loops in the brain.

Several research threads open, of which we discuss a few important ones to spur discussion within the community. The success of any model hinges on data. Therefore, the first most important question is (1) *how to effectively capture and curate interactive experiences between a human and a robot.* Autonomous-driving consortia have released large multi-agent datasets, V2X-SIM [75], OPV2V [139], DAIR-V2X [142], that pair synchronized sensor logs with ground-truth maps. Yet, only a few capture the bidirectional dynamics of a human-robot interaction [90, 18]. New corpora must record the closed-loop triad ⟨robot state, environment state, human internal state proxies⟩ at millisecond scale, and must annotate not only "what was seen" but "what was useful." Wearable inertial units, body-mounted cameras, and privacy-preserving eye trackers are promising instrumentation for raw data collection, with emerging HRI datasets such as PARTNR [18] going in this direction; the open challenge is to curate and synchronize more meta-level features such as humans' interactive experience, cognitive state, comfort, fatigue, and trust levels. Another consideration is the difference in timescales of the functioning of humans and robots. (2) *How can we measure true collaboration when human intent and actions unfold far faster than robot policy updates?* Simultaneous sampling blurs causality: the human adapts before the robot moves. A practical fix, used in PARTNR, is dual-rate logging: first, run the task online with real robot latency, then replay the same robot actions time-warped to the human clock. However, this still assumes the robot would choose the same actions if it could react earlier and ignores real-world factors, sensor noise, network jitter, actuator limits, and privacy-driven sensing gaps, that reshape both partners' decisions; capturing true collaboration will require hardware-level traces that synchronise heterogeneous clocks and explicitly label moments

where reduced latency would have changed either agent's plan. Further, while this remedy is feasible for largely autonomous robots whose decisions are locally self-consistent, it is not useful for wearable robots whose perception-action cycle is tightly coupled to the human.

4.2 Module 2: Teamwork Model for Collaborative Actions

Human-robot cooperation hinges on how well each partner can keep track of the other's current internal state and fold that estimate into its own action choices. This requires modeling the human and robot as two co-agents collaborating with imperfect information. The robot should model human's condition (e.g., intent, fatigue, affective state, comfort thresholds) as a latent state to be inferred and updated continuously, not as a fixed input [4]. For instance, in shared-control teleoperation or exoskeleton use, the operator or wearer adapts to the device's actions as much as vice versa. The robot should therefore, use all available cues (e.g., EMG or motion patterns, visual observations, spoken instructions) to infer the user's goals and actions. This enables coordination. One example is presented in [87] where a prosthesis controller uses multimodal sensor input to anticipate the target object the user would like to grasp, estimate its size and shape, and preshape the device accordingly, all of this while tracking and reacting to user movements (e.g., a change in the approach trajectory).

Parallels can be drawn in cognitive science, where joint action explores how individuals coordinate tasks by internally representing each other's goals and actions [111]. The brain also employs partial models of another person's internal states in collaborative tasks, often referred to as *shared intentionality* [11]. When humans collaborate, they engage in *shared intentionality*, and mutual understanding of each other's intentions, states, and possible actions [11]. Neuroscientific research further reveals that humans utilize a form of theory of mind [6], allowing one to infer another's mental states, thereby enabling synchronization in activities such as dancing, carrying a table together, or passing objects. These mechanisms underpin our ability to anticipate partners' behavior, rapidly adapt to unexpected changes, and maintain coordinated trajectories.

Robot foundation models can replicate the same skill set through three complementary sub-functions: 1) *Intent inference*: although the robot (e.g., a leg prosthesis) cannot directly "read" the user's mind, it can decode immediate goals and motor patterns from EMG envelopes, limb dynamics, or vision sensors placed on the prosthesis [69]. 2) *Belief-state maintenance*: by building on the theory of mind [6], the robot keeps a running posterior over latent human variables. For the bionic limbs, this includes preferred torque bands, comfort envelopes, onset of fatigue, and corrects the posterior whenever sensed outcomes diverge from its own prediction. However, a specific challenge is how to implement such prediction and model correction in a safe manner (e.g., a wrong prediction in a lower limb prosthesis can lead to falling). 3) *Proposal refinement*: The teamwork layer filters high-level action proposals from the sensing module through user-centric constraints (e.g., joint stress) and situational collaboration tactics (e.g., aligning push-off timing with the user's weight shift).

Several research avenues open while incorporating a teamwork mechanism into robot foundation models. (1) Since keeping a theory-of-mind of the partner (human) is important for human-robot collaboration, one may ask *how can a robot do this data-efficiently?*. A promising route could be to pre-train in large simulated multi-agent worlds [93, 125] and then meta-learn a fast adapter that re-weights the prior online. (2) From the model training perspective, a design choice would be *whether to propagate the gradients from teamwork objectives back into perception*. Multi-task reinforcement learning already shows that a shared encoder feeding multiple task-specific critics promotes transfer without catastrophic forgetting [137]. Extending that idea, recent world-model-based agents expose intermediate latent states to an explanation head, forcing the encoder to stay semantically aligned with downstream reasoning [114]. Applying the same principle across the sensing-teamwork boundary could yield perceptual features intrinsically shaped for cooperative inference.

4.3 Module 3: Predictive World Belief Model for Anticipation and Planning

Where the teamwork module concentrates on the *present* dyad state, the predictive world belief model looks *ahead*. Its purpose is to simulate how the human, robot, and environment are likely to co-evolve over the next few hundred milliseconds. Human movements usually unfold faster than a robot can sense, process, and react. To keep up with the required pace of actions, a human-interactive robot must do more than react to what is; it must reason about what will soon be true for both itself and its partner. The Predictive World Belief Model supplies this anticipatory layer by learning an internal, probabilistic forward model of the dyad-plus-environment. By rolling forward a learned dynamics

model, the robot gains a distribution over future latent variables such as user intent, muscle fatigue trajectories, object affordances (e.g., for hand prostheses), or terrain parameters (e.g., for lower-limb prostheses and exoskeletons). These forecasts are fed back into planning so that the controller can bias actions toward outcomes that will remain feasible and safe once the human commits.

Neuroscience offers parallels: Research in motor neuroscience underscores the role of *internal forward models*, which predict the sensory outcomes of motor commands, adjusting subsequent behavior in anticipation of future states. [134, 64]. Predictive coding theories go further, proposing that the brain continuously attempts to minimize prediction errors by updating these models [46]. Translating this to HRI, the robot should treat the human as a stochastic dynamical subsystem whose future states can be inferred from present cues, much as the brain treats another person's intent within its predictive coding hierarchy.

Integrating predictive world models into human-interactive robotic foundation models opens several important research directions. (1) *Handling uncertainty and non-stationarity:* Human behavior can shift abruptly due to internal state changes such as fatigue or distraction. Predictive models must detect such distributional shifts early and re-plan accordingly. Online Bayesian change point detection applied to physiological and kinematic signals can identify fatigue onset before performance declines [17, 92]. Embedding these detectors into belief update mechanisms would enable adaptive modulation of assistive control. Deviations of world model predictions from real outcomes can also signal changes in the user strategy or environment characteristics, prompting model updates. Fast Bayesian belief updates combined with meta-learned priors can facilitate rapid adaptation with minimal interaction data [63, 55, 105]. (2) *Utilizing predictions for control:* A key design question is how controllers should leverage predictions. One option is to continuously adjust control parameters to minimize predicted risk, akin to a real-time safety monitor. Another is to use predictions for internal look-ahead planning, selecting actions based on simulated outcomes. These strategies may be implemented via model predictive control (MPC), provided predictive models are computationally efficient [107]. Further research is needed to assess the viability of learning-based models in MPC or game-theoretic planners for human-robot teams. (3) *Temporal planning horizons:* As anticipatory control is the main functionality that the predictive world model satisfies, another consideration is the planning horizon. Long-term objectives (e.g., energy conservation) often conflict with the need for rapid reflexes. Hierarchical belief stacks, which maintain temporally layered models, offer a promising solution [49], potentially enabling slow planners to track latent states like fatigue, while fast controllers handle immediate dynamics. (4) *Incorporating prior knowledge:* Predictive models may benefit from encoding prior world knowledge. Physics-informed neural networks embed physical laws into the training objective, enabling more robust generalization under limited data [77]. However, their application in human-robot interaction remains rather underexplored [141, 140]. Addressing these challenges is necessary to transform a forward predictive model into the anticipatory core of next-generation collaborative and assistive robots that not only react to the user's current intent but routinely predict it and plan future actions accordingly.

4.4 Module 4: Memory & Feedback for Refinement

Continuous interactions with humans allow the robot to learn and improve over time. This requires the robot to store previous interactions, user preferences, and feedback in its memory. A dedicated Memory & Feedback mechanism gives the system a place to accumulate user-specific knowledge and a pathway for feedback signals to reshape future control. This enables the device to be more fluent, energy-efficient, and trustworthy as it interacts more with the user.

Neuroscience emphasizes how synaptic plasticity drives *long-term* changes in behavior through feedback-driven processes, including dopamine-mediated error- and reward-related reinforcement signals [48, 27]. For example, repetitive practice consolidates motor memories that lead to progressive refinement of motor representations in the brain and improved task efficiency. The Memory & Feedback mechanism may adopt the same architecture: it stores traces of past state-action-outcome tuples, estimates a scalar reward (metabolic cost, user comfort), and adjusts internal parameters so that future actions shift probability mass toward higher reward.

The *Memory* and *Feedback Mechanism* should offer certain critical functionalities such as 1) *Long-term preference storage*: The device (e.g., a leg prosthesis) stores user-specific torque settings, comfort ranges, and typical walking patterns [35], similar to how repeated exposure to a task solidifies neural pathways in motor learning. Similarly, a hand prosthesis could remember how the user prefers to

grasp particular objects in their home. 2) *Reinforcement-based updates*: The device can query the user explicitly ("Is this stiffness comfortable?") when the adjustment function is activated or automatically evaluate signals of discomfort, in the background, by adjusting internal parameters to optimize an objective function (e.g., minimal metabolic cost, subjective comfort). 3) *Semantic memory bank*: Language-based preferences ("I like a softer ankle when walking on grass") can be retained as textual or embedding-based knowledge, during the initial setup phase, allowing the device to recall and apply them in the future.

Designing an effective memory and feedback mechanism for human-robot interaction presents several challenges. (1) Given constraints on memory and computation, particularly in real-time settings, it is critical to determine *what information should be retained*. Storing all state-action pairs is impractical; thus, approaches such as map-based storage [53] and value-aware compression [108] can prioritize novel or high-salience experiences, such as those linked to user discomfort. Retrieval-Augmented Embodied Agents (RAEA) [144] implement this by retrieving relevant strategies from external memory and integrating them into learning via a generator module. (2) While the ability to store and leverage prior interactions is powerful, it raises the challenge of *incorporating new knowledge without catastrophic forgetting*. Continual learning techniques, including progressive networks and episodic memory, help preserve existing capabilities while adapting to user-specific data [97]. However, applying these methods to human-robot interaction remains an open problem [5]. (3) A further concern is *data privacy*. Robots must avoid leaking sensitive personal information. Federated learning approaches, such as FedHIP [15], demonstrate that encrypted gradient sharing can support intention prediction without exposing raw data. Whether such privacy-preserving mechanisms can be generalized to multimodal, interactive control scenarios remains to be seen. Addressing these issues is essential for evolving robots from generic assistive devices into personalized, adaptive partners continually shaped by the user's feedback and changing needs..

5 Prioritizing Function, Not Form

A monolithic, end-to-end trained foundation model that ingests the full stream of perceptual, physiological, and interaction history, and directly outputs robot actions, can, in principle, internalize the very competencies we advocate for. This reflects the prevailing philosophy behind many foundation models: with sufficient data and a powerful architecture (e.g., large transformers), such models *might* implicitly learn to integrate human feedback, predict outcomes, and adapt to evolving tasks.

However, emerging research suggests that monolithic architectures benefit from augmentation with explicit auxiliary modules, such as, memory mechanisms for tracking user preferences, or, latent models of co-agents [5, 46, 82]. This trend aligns with our position: rather than assuming that scale alone will yield all necessary competencies, we argue for embedding key multi-agent capabilities, such as co-adaptation, predictive modeling, and user state inference, into the model design. These enhancements echo cognitive and neuroscientific insights into memory, anticipation, and adaptation, which are critical for effective human-robot co-adaptation. Rather than prescribing a specific architecture, fully modular, monolithic, or hybrid, we emphasize the importance of embedding key capabilities (e.g., multi-agent collaboration, user modeling, predictive reasoning) into future robotic foundation models. The optimal design will likely depend on application-specific tradeoffs in adaptability, trainability, and interpretability. As the field rapidly evolves, our goal is to steer foundational research toward integrating multi-agent principles early, ensuring that future models are well-equipped for interactive, personalized, and robust human-robot collaboration.

6 Safety, Ethical, and Regulatory Perspectives

Safety and Interpretability. Human-interactive robots must incorporate safety layers and be explainable. A low-level reflex controller, analogous to spinal reflexes [131], can enforce hard constraints: for instance, immediate shutdown or torque limits if joint angles become unsafe. Above this, higher-level policies should be transparent so that clinicians or users can understand why a particular action was chosen. Techniques from human-aware AI (e.g., generating natural-language justifications) and explainable AI may help build trust [51, 70]. Extensive validation, especially for medical devices, must accompany any deployment.

Privacy, Data Ownership, and Bias Mitigation. Multi-agent systems often require extensive collection and processing of physiological and contextual data, increasing privacy concerns. Implementing robust data protection measures, such as on-device processing [138], federated learning [143, 112], and anonymized data management [118, 23], is essential to safeguard user information. Additionally, large-scale models can inadvertently perpetuate biases present in their training data [9]. In healthcare and assistive contexts, such biases could lead to unequal performance across different user groups. Thus, rigorous data curation, bias mitigation strategies, and domain-specific fine-tuning are imperative to ensure equitable and unbiased system performance.

Medical Accountability and Clinical Evidence. Adopting multi-agent foundation models in robotics, particularly within medical and assistive technologies, necessitates stringent compliance with regulatory standards [106, 16]. The integration of multiple autonomous agents introduces complexities in risk assessment and accountability [89]. It is imperative to conduct comprehensive clinical trials and gather substantial evidence to validate the safety and efficacy of these AI-driven control strategies. Ensuring minimal risks of adverse events, such as falls or device malfunctions, and demonstrating reliable performance across diverse user populations is critical for regulatory approval and widespread adoption [9]. However, an advantage of multi-agent models is that they could integrate risk assessment and mitigation as an intrinsic function based on the comprehensive insight into the state of the human user and the environment, which such models have.

7 Conclusion and Outlook

In this paper, we argued that the prevailing single-agent paradigm in robotic foundation models is fundamentally inadequate for scenarios requiring rich, continuous human-robot interaction. We proposed a shift toward an interactive multi-agent framework, where both the human and the robot are treated as dynamically adapting agents. Drawing on insights from neuroscience and cognitive science, we outlined four key functionalities, multimodal collaborative sensing, teamwork modeling, predictive world belief models, and memory/feedback mechanisms, needed in robotic foundation models to realize this vision. While we grounded our arguments mainly in wearable and assistive robotics, the principles extend broadly to any context where real-time co-adaptation is critical.

To advance interactive intelligence, the community must develop benchmarks and datasets focused on human-robot co-adaptation. Just as ImageNet and COCO catalyzed breakthroughs in computer vision, large-scale datasets capturing human-robot interaction can be similarly transformative. The Open X-Embodiment dataset [26], which includes over one million robot trajectories across platforms, exemplifies this potential. We advocate for similar open datasets enriched with multi-modal human cues. Standardized tasks involving human partners, such as collaborative object manipulation or adaptive gait, would enable consistent benchmarking. Additionally, shared simulation environments with embodied virtual humans could accelerate development through rapid prototyping and sim-to-real transfer. In practice, widespread progress will depend on: (1) *Standardized benchmarks*: Similar to the large-scale robotics benchmarks in manipulation [88, 84, 65], new testbeds are required for human-interactive robots, such as robotic prostheses and exoskeletons. These testbeds should capture real-world complexity such as irregular terrains, evolving user states (e.g., fatigue), and long-term usage scenarios [88, 84, 65]. (2) *Open-source ecosystems*: encouraging shared datasets and sharing pre-trained models can significantly expedite research, as they have in computer vision and NLP. Large-scale corpora for wearable robotics would play a foundational role in advancing interactive models. (3) *Clinical partnerships* with therapists, clinicians, and end-users are essential to ensure that robotic systems meet real-world functional goals, such as reducing fall risk, improving metabolic efficiency, and enhancing subjective comfort.

Finally, human-in-the-loop learning should be explored. Instead of passively learning from fixed logs, future systems could actively query users for feedback or clarification. For example, a robot might ask "Did I interpret your intent correctly?" in real time, using the answer to update its model. In wearable robotics, where speaking to a robot is not that common, such communication could be implemented through tactile cues and usual command interfaces (e.g., a specific muscle pattern could indicate to the system that the last estimation was wrong). Integrating such interaction schemes with foundation models could greatly improve personalization and robustness.

In summary, we argue that next-generation robot models must move beyond isolated, single-shot policies. By adopting a multi-agent perspective, drawing on cognitive and neuroscientific insights,

robots can achieve more flexible, personalized, and anticipatory behaviors. We hope this position paper sparks discussion on designing and training interactive foundation models that ultimately enrich human-robot co-adaptation and generalization.

References

[1] Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of eeg-based brain–computer interface paradigms. *Journal of neural engineering*, 16(1):011001, 2019.

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[3] Stefano V Albrecht, Filippos Christianos, and Lukas Schäfer. *Multi-agent reinforcement learning: Foundations and modern approaches*. MIT Press, 2024.

[4] Stefano V Albrecht and Peter Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95, 2018.

[5] Ali Ayub, Zachary De Francesco, Patrick Holthaus, Chrystopher L Nehaniv, and Kerstin Dautenhahn. Continual learning through human-robot interaction: Human perceptions of a continual learning robot in repeated interactions. *International Journal of Social Robotics*, pages 1–20, 2025.

[6] Simon Baron-Cohen, Howard Ring, John Moriarty, Bettina Schmitz, Durval Costa, and Peter Ell. Recognition of mental state terms. *British Journal of Psychiatry*, 165(5):640–649, 1994.

[7] Samuel Barrett. *Making friends on the fly: advances in ad hoc teamwork*, volume 603. Springer, 2015.

[8] Romain Baud, Ali Reza Manzoori, Auke Ijspeert, and Mohamed Bouri. Review of control strategies for lower-limb exoskeletons to assist gait. *Journal of neuroengineering and rehabilitation*, 18:1–34, 2021.

[9] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[10] T Kevin Best, Cara Gonzalez Welker, Elliott J Rouse, and Robert D Gregg. Data-driven variable impedance control of a powered knee–ankle prosthesis for adaptive speed and incline walking. *IEEE Transactions on Robotics*, 39(3):2151–2169, 2023.

[11] Michael E Bratman. Shared cooperative activity. *The philosophical review*, 101(2):327–341, 1992.

[12] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.

[13] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

[14] Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. *NeurIPS*, 2020.

[15] Jiannan Cai, Zhidong Gao, Yuanxiong Guo, Bastian Wibranek, and Shuai Li. Fedhip: Federated learning for privacy-preserving human intention prediction in human-robot collaborative assembly tasks. *Advanced Engineering Informatics*, 60:102411, 2024.

[16] Praminda Caleb-Solly, Chris Harper, and Sanja Dogramadzi. Standards and regulations for physically assistive robots. In *2021 IEEE international conference on intelligence and safety for robotics (ISR)*, pages 259–263. IEEE, 2021.

[17] Saahil Chand, Hao Zheng, and Yuqian Lu. A vision-enabled fatigue-sensitive human digital twin towards human-centric human-robot collaboration. *Journal of Manufacturing Systems*, 77:432–445, 2024.

[18] Matthew Chang, Gunjan Chhablani, Alexander Clegg, Mikael Dallaire Cote, Ruta Desai, Michal Hlavac, Vladimir Karashchuk, Jacob Krantz, Roozbeh Mottaghi, Priyam Parashar, et al. Partnr: A benchmark for planning and reasoning in embodied multi-agent tasks. *arXiv preprint arXiv:2411.00081*, 2024.

[19] Harrison Chase and LangChain Contributors. Langchain: Build applications with llms through composability. `https://github.com/langchain-ai/langchain`, 2023. Accessed: May 17, 2025.

[20] Guangyao Chen, Siwei Dong, Yu Shu, Ge Zhang, Jaward Sesay, Börje F Karlsson, Jie Fu, and Yemin Shi. Autoagents: A framework for automatic agent generation. *arXiv preprint arXiv:2309.17288*, 2023.

[21] Xiaodong Chen, Gregory C DeAngelis, and Dora E Angelaki. Flexible egocentric and allocentric representations of heading signals in parietal cortex. *Proceedings of the National Academy of Sciences*, 115(14):E3305–E3312, 2018.

[22] Federico Chiariotti, Pranav Mamidanna, Suraj Suman, Čedomir Stefanović, Dario Farina, Petar Popovski, and Strahinja Došen. The future of bionic limbs: The untapped synergy of signal processing, control, and wireless connectivity. *IEEE Signal Processing Magazine*, 41(4):58–75, 2024.

[23] Olivia Choudhury, Aris Gkoulalas-Divanis, Theodoros Salonidis, Issa Sylla, Yoonyoung Park, Grace Hsu, and Amar Das. Anonymizing data for privacy-preserving federated learning. *arXiv preprint arXiv:2002.09096*, 2020.

[24] Andrea Cimolato, Josephus JM Driessen, Leonardo S Mattos, Elena De Momi, Matteo Laffranchi, and Lorenzo De Michieli. Emg-driven control in lower limb prostheses: A topic-based systematic review. *Journal of NeuroEngineering and Rehabilitation*, 19(1):43, 2022.

[25] Hallie Clark and Jing Feng. Semi-autonomous vehicles: Examining driver performance during the take-over. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 781–785. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[26] Open X-Embodiment Collaboration, Abby O'Neill, Abdul Rehman, Abhinav Gupta, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, Albert Tung, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anchit Gupta, Andrew Wang, Andrey Kolobov, Anikait Singh, Animesh Garg, Aniruddha Kembhavi, Annie Xie, Anthony Brohan, Antonin Raffin, Archit Sharma, Arefeh Yavary, Arhan Jain, Ashwin Balakrishna, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Blake Wulfe, Brian Ichter, Cewu Lu, Charles Xu, Charlotte Le, Chelsea Finn, Chen Wang, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Christopher Agia, Chuer Pan, Chuyuan Fu, Coline Devin, Danfei Xu, Daniel Morton, Danny Driess, Daphne Chen, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dinesh Jayaraman, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Ethan Foster, Fangchen Liu, Federico Ceola, Fei Xia, Feiyu Zhao, Felipe Vieira Frujeri, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Gilbert Feng, Giulio Schiavi, Glen Berseth, Gregory Kahn, Guangwen Yang, Guanzhi Wang, Hao Su, Hao-Shu Fang, Haochen Shi, Henghui Bao, Heni Ben Amor, Henrik I Christensen, Hiroki Furuta, Homanga Bharadhwaj, Homer Walke, Hongjie Fang, Huy Ha, Igor Mordatch, Ilija Radosavovic, Isabel Leal, Jacky Liang, Jad Abou-Chakra, Jaehyung Kim, Jaimyn Drake, Jan Peters, Jan Schneider, Jasmine Hsu, Jay Vakil, Jeannette Bohg, Jeffrey Bingham, Jeffrey Wu, Jensen Gao, Jiaheng Hu, Jiajun Wu, Jialin Wu, Jiankai Sun, Jianlan Luo, Jiayuan Gu, Jie Tan, Jihoon Oh, Jimmy Wu, Jingpei Lu, Jingyun

Yang, Jitendra Malik, João Silvério, Joey Hejna, Jonathan Booher, Jonathan Tompson, Jonathan Yang, Jordi Salvador, Joseph J. Lim, Junhyek Han, Kaiyuan Wang, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin Black, Kevin Lin, Kevin Zhang, Kiana Ehsani, Kiran Lekkala, Kirsty Ellis, Krishan Rana, Krishnan Srinivasan, Kuan Fang, Kunal Pratap Singh, Kuo-Hao Zeng, Kyle Hatch, Kyle Hsu, Laurent Itti, Lawrence Yunliang Chen, Lerrel Pinto, Li Fei-Fei, Liam Tan, Linxi "Jim" Fan, Lionel Ott, Lisa Lee, Luca Weihs, Magnum Chen, Marion Lepert, Marius Memmel, Masayoshi Tomizuka, Masha Itkina, Mateo Guaman Castro, Max Spero, Maximilian Du, Michael Ahn, Michael C. Yip, Mingtong Zhang, Mingyu Ding, Minho Heo, Mohan Kumar Srirama, Mohit Sharma, Moo Jin Kim, Muhammad Zubair Irshad, Naoaki Kanazawa, Nicklas Hansen, Nicolas Heess, Nikhil J Joshi, Niko Suenderhauf, Ning Liu, Norman Di Palo, Nur Muhammad Mahi Shafiullah, Oier Mees, Oliver Kroemer, Osbert Bastani, Pannag R Sanketi, Patrick "Tree" Miller, Patrick Yin, Paul Wohlhart, Peng Xu, Peter David Fagan, Peter Mitrano, Pierre Sermanet, Pieter Abbeel, Priya Sundaresan, Qiuyu Chen, Quan Vuong, Rafael Rafailov, Ran Tian, Ria Doshi, Roberto Mart'in-Mart'in, Rohan Baijal, Rosario Scalise, Rose Hendrix, Roy Lin, Runjia Qian, Ruohan Zhang, Russell Mendonca, Rutav Shah, Ryan Hoque, Ryan Julian, Samuel Bustamante, Sean Kirmani, Sergey Levine, Shan Lin, Sherry Moore, Shikhar Bahl, Shivin Dass, Shubham Sonawani, Shubham Tulsiani, Shuran Song, Sichun Xu, Siddhant Haldar, Siddharth Karamcheti, Simeon Adebola, Simon Guist, Soroush Nasiriany, Stefan Schaal, Stefan Welker, Stephen Tian, Subramanian Ramamoorthy, Sudeep Dasari, Suneel Belkhale, Sungjae Park, Suraj Nair, Suvir Mirchandani, Takayuki Osa, Tanmay Gupta, Tatsuya Harada, Tatsuya Matsushima, Ted Xiao, Thomas Kollar, Tianhe Yu, Tianli Ding, Todor Davchev, Tony Z. Zhao, Travis Armstrong, Trevor Darrell, Trinity Chung, Vidhi Jain, Vikash Kumar, Vincent Vanhoucke, Vitor Guizilini, Wei Zhan, Wenxuan Zhou, Wolfram Burgard, Xi Chen, Xiangyu Chen, Xiaolong Wang, Xinghao Zhu, Xinyang Geng, Xiyuan Liu, Xu Liangwei, Xuanlin Li, Yansong Pang, Yao Lu, Yecheng Jason Ma, Yejin Kim, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Yilin Wu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yongqiang Dou, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yue Cao, Yueh-Hua Wu, Yujin Tang, Yuke Zhu, Yunchu Zhang, Yunfan Jiang, Yunshuang Li, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zehan Ma, Zhuo Xu, Zichen Jeff Cui, Zichen Zhang, Zipeng Fu, and Zipeng Lin. Open X-Embodiment: Robotic learning datasets and RT-X models. `https://arxiv.org/abs/2310.08864`, 2023.

[27] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.

[28] Susan L Denham and István Winkler. Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience*, 51(5):1151–1160, 2020.

[29] Sharmita Dey. *Learning-based Biomimetic Strategies for Developing Control Schemes for Lower Extremity Rehabilitation Robotic Devices*. PhD thesis, Georg-August-Universität Göttingen, 2023.

[30] Sharmita Dey, Sabri Boughorbel, and Arndt F Schilling. Learning a shared model for motorized prosthetic joints to predict ankle-joint motion. *arXiv preprint arXiv:2111.07419*, 2021.

[31] Sharmita Dey, Niklas De Schultz, and Arndt F Schilling. Why hard code the bionic limbs when they can learn from humans? In *2023 International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE, 2023.

[32] Sharmita Dey, Mahdy Eslamy, Takashi Yoshida, Michael Ernst, Thomas Schmalz, and ArndtF Schilling. A support vector regression approach for continuous prediction of ankle angle and moment during walking: An implication for developing a control strategy for active ankle prostheses. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, pages 727–733. IEEE, 2019.

[33] Sharmita Dey, David Fan, Robin Schmid, Anushri Dixit, Kyohei Otsu, Thomas Touma, Arndt F Schilling, and Ali-Akbar Agha-Mohammadi. Prepare: Predictive proprioception for agile failure event detection in robotic exploration of extreme terrains. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4338–4343. IEEE, 2022.

[34] Sharmita Dey and Sarath R Nair. Enhancing joint motion prediction for individuals with limb loss through model reprogramming. *arXiv preprint arXiv:2403.06569*, 2024.

[35] Sharmita Dey, Benjamin Paassen, Sarath Ravindran Nair, Sabri Boughorbel, and Arndt F Schilling. Continual learning from simulated interactions via multitask prospective rehearsal for bionic limb behavior modeling. *arXiv preprint arXiv:2405.01114*, 2024.

[36] Sharmita Dey and Sarath Ravindran Nair. Remap: Neural model reprogramming with network inversion and retrieval-augmented mapping for adaptive motion forecasting. *Advances in Neural Information Processing Systems*, 37:25195–25227, 2024.

[37] Sharmita Dey and Arndt F Schilling. Data-driven gait-predictive model for anticipatory prosthesis control. In *2022 International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE, 2022.

[38] Sharmita Dey and Arndt F Schilling. A function approximator model for robust online foot angle trajectory prediction using a single imu sensor: Implication for controlling active prosthetic feet. *IEEE Transactions on Industrial Informatics*, 19(2):1467–1475, 2022.

[39] Sharmita Dey, Takashi Yoshida, Michael Ernst, Thomas Schmalz, and Arndt F Schilling. A random forest approach for continuous prediction of joint angles and moments during walking: An implication for controlling active knee-ankle prostheses/orthoses. In *2019 IEEE International conference on Cyborg and bionic systems (CBS)*, pages 66–71. IEEE, 2019.

[40] Sharmita Dey, Takashi Yoshida, Robert H Foerster, Michael Ernst, Thomas Schmalz, Rodrigo M Carnier, and Arndt F Schilling. A hybrid approach for dynamically training a torque prediction model for devising a human-machine interface control strategy. *arXiv preprint arXiv:2110.03085*, 2021.

[41] Sharmita Dey, Takashi Yoshida, Robert H Foerster, Michael Ernst, Thomas Schmalz, and Arndt F Schilling. Continuous prediction of joint angular positions and moments: A potential control strategy for active knee-ankle prostheses. *IEEE Transactions on Medical Robotics and Bionics*, 2(3):347–355, 2020.

[42] Sharmita Dey, Takashi Yoshida, and Arndt F Schilling. Feasibility of training a random forest model with incomplete user-specific data for devising a control strategy for active biomimetic ankle. *Frontiers in Bioengineering and Biotechnology*, 8:855, 2020.

[43] Anushri Dixit, David D Fan, Kyohei Otsu, Sharmita Dey, Ali-Akbar Agha-Mohammadi, and Joel Burdick. Step: Stochastic traversability evaluation and planning for risk-aware navigation; results from the darpa subterranean challenge. *Field Robotics*, 4:182–210, 2024.

[44] John P Donoghue. Bridging the brain to the world: a perspective on neural interface systems. *Neuron*, 60(3):511–521, 2008.

[45] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.

[46] Karl Friston. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, 13(7):293–301, 2009.

[47] Rachel Gehlhar, Maegan Tucker, Aaron J Young, and Aaron D Ames. A review of current state-of-the-art control methods for lower-limb powered prostheses. *Annual Reviews in Control*, 55:142–164, 2023.

[48] Wulfram Gerstner and Werner M Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

[49] Christian Gumbsch, Noor Sajid, Georg Martius, and Martin V Butz. Learning hierarchical world models with adaptive temporal abstractions from discrete latent dynamics. In *The Twelfth International Conference on Learning Representations*, 2023.

[50] Weichao Guo, Wei Xu, Yanchao Zhao, Xu Shi, Xinjun Sheng, and Xiangyang Zhu. Toward human-in-the-loop shared control for upper-limb prostheses: a systematic analysis of state-of-the-art technologies. *IEEE transactions on Medical Robotics and Bionics*, 5(3):563–579, 2023.

[51] Balint Gyevnar, Cheng Wang, Christopher G Lucas, Shay B Cohen, and Stefano V Albrecht. Causal explanations for sequential decision-making in multi-agent systems. *arXiv preprint arXiv:2302.10809*, 2023.

[52] David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2(3), 2018.

[53] Muhammad Burhan Hafez, Tilman Immisch, Tom Weber, and Stefan Wermter. Map-based experience replay: a memory-efficient solution to catastrophic forgetting in reinforcement learning. *Frontiers in Neurorobotics*, 17:1127642, 2023.

[54] Dongge Han, Trevor McInroe, Adam Jelley, Stefano V. Albrecht, Peter Bell, and Amos Storkey. Llm-personalize: Aligning llm planners with human preferences via reinforced self-training for housekeeping robots. In *International Conference on Computational Linguistics*, 2024.

[55] James Harrison, Apoorva Sharma, and Marco Pavone. Meta-learning priors for efficient online bayesian regression. In *International Workshop on the Algorithmic Foundations of Robotics*, pages 318–337. Springer, 2018.

[56] Guy Hoffman, Tapomayukh Bhattacharjee, and Stefanos Nikolaidis. Inferring human intent and predicting human action in human–robot collaboration. *Annual Review of Control, Robotics, and Autonomous Systems*, 7, 2023.

[57] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022.

[58] Yue Hu, Xianghe Pang, Xiaoqi Qin, Yonina C Eldar, Siheng Chen, Ping Zhang, and Wenjun Zhang. Pragmatic communication in multi-agent collaborative perception. *arXiv preprint arXiv:2401.12694*, 2024.

[59] Tao Huang, Jianan Liu, Xi Zhou, Dinh C Nguyen, Mostafa Rahimi Azghadi, Yuxuan Xia, Qing-Long Han, and Sumei Sun. V2x cooperative perception for autonomous driving: Recent advances and challenges. *arXiv preprint arXiv:2310.03525*, 2023.

[60] Silvio Ionta, Lukas Heydrich, Bigna Lenggenhager, Michael Mouthon, Eleonora Fornari, Dominique Chapuis, Roger Gassert, and Olaf Blanke. Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron*, 70(2):363–374, 2011.

[61] Andrew Jackson and Jonas B Zimmermann. Neural interfaces for the brain and spinal cord—restoring motor function. *Nature Reviews Neurology*, 8(12):690–699, 2012.

[62] Ning Jiang, Strahinja Dosen, Klaus-Robert Muller, and Dario Farina. Myoelectric control of artificial limbs—is there a need to change focus?[in the spotlight]. *IEEE Signal Processing Magazine*, 29(5):152–150, 2012.

[63] Rituraj Kaushik, Timothée Anne, and Jean-Baptiste Mouret. Fast online adaptation in robotics through meta-learning embeddings of simulated priors. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5269–5276. IEEE, 2020.

[64] Mitsuo Kawato. Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, 9(6):718–727, 1999.

[65] Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, et al. Droid: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

[66] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[67] Elsa Andrea Kirchner and Judith Bütefür. Towards bidirectional and coadaptive robotic exoskeletons for neuromotor rehabilitation and assisted daily living: a review. *Current Robotics Reports*, 3(2):21–32, 2022.

[68] Jonathan Kofman, Xianghai Wu, Timothy J Luu, and Siddharth Verma. Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE transactions on industrial electronics*, 52(5):1206–1219, 2005.

[69] Nili E Krausz and Levi J Hargrove. A survey of teleceptive sensing for wearable assistive robotic devices. *Sensors*, 19(23):5238, 2019.

[70] Anton Kuznietsov, Balint Gyevnar, Cheng Wang, Steven Peters, and Stefano V Albrecht. Explainable ai for safe and trustworthy autonomous driving: a systematic review. *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[71] Yann LeCun. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1):1–62, 2022.

[72] Huao Li, Tianwei Ni, Siddharth Agrawal, Fan Jia, Suhas Raja, Yikang Gui, Dana Hughes, Michael Lewis, and Katia Sycara. Individualized mutual adaptation in human-agent teams. *IEEE Transactions on Human-Machine Systems*, 51(6):706–714, 2021.

[73] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023.

[74] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR, 2022.

[75] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters*, 7(4):10914–10921, 2022.

[76] S Lichiardopol. A survey on teleoperation. 2007.

[77] Jingyue Liu, Pablo Borja, and Cosimo Della Santina. Physics-informed neural networks to model and control robots: A theoretical and experimental investigation. *Advanced Intelligent Systems*, 6(5):2300385, 2024.

[78] Yen-Cheng Liu, Junjiao Tian, Nathaniel Glaser, and Zsolt Kira. When2com: Multi-agent perception via communication graph grouping. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4106–4115, 2020.

[79] Ho Shing Lo and Sheng Quan Xie. Exoskeleton robots for upper-limb rehabilitation: State of the art and future prospects. *Medical engineering & physics*, 34(3):261–268, 2012.

[80] Dennis J McFarland and Jonathan R Wolpaw. Eeg-based brain–computer interfaces. *current opinion in Biomedical Engineering*, 4:194–200, 2017.

[81] Francisco S Melo and Alberto Sardinha. Ad hoc teamwork by learning teammates' task. *Autonomous Agents and Multi-Agent Systems*, 30:175–219, 2016.

[82] Beren Millidge, Anil Seth, and Christopher L Buckley. Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979*, 2021.

[83] Reuth Mirsky, Ignacio Carlucho, Arrasy Rahman, Elliot Fosong, William Macke, Mohan Sridharan, Peter Stone, and Stefano V Albrecht. A survey of ad hoc teamwork research. In *European conference on multi-agent systems*, pages 275–293. Springer, 2022.

[84] Chaitanya Mitash, Fan Wang, Shiyang Lu, Vikedo Terhuja, Tyler Garaas, Felipe Polido, and Manikantan Nambi. Armbench: An object-centric benchmark dataset for robotic manipulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9132–9139. IEEE, 2023.

[85] Franco Molteni, Giulio Gasperini, Giovanni Cannaviello, and Eleonora Guanziroli. Exoskeleton and end-effector robots for upper and lower limbs rehabilitation: narrative review. *PM&R*, 10(9):S174–S188, 2018.

[86] Benjamin Morrell, Kyohei Otsu, Ali Agha, David D Fan, Sung-Kyun Kim, Muhammad Fadhil Ginting, Xianmei Lei, Jeffrey Edlund, Seyed Fakoorian, Amanda Bouman, et al. An addendum to nebula: Towards extending team costar's solution to larger scale environments. *IEEE Transactions on Field Robotics*, 2024.

[87] Jeremy Mouchoux, Stefano Carisi, Strahinja Dosen, Dario Farina, Arndt F Schilling, and Marko Markovic. Artificial perception and semiautonomous control in myoelectric hand prostheses increases performance and decreases effort. *IEEE Transactions on Robotics*, 37(4):1298–1312, 2021.

[88] Tongzhou Mu, Zhan Ling, Fanbo Xiang, Derek Yang, Xuanlin Li, Stone Tao, Zhiao Huang, Zhiwei Jia, and Hao Su. Maniskill: Generalizable manipulation skill benchmark with large-scale demonstrations. *arXiv preprint arXiv:2107.14483*, 2021.

[89] Amirreza Naseri, Ming Liu, I-Chieh Lee, Wentao Liu, and He Huang. Characterizing prosthesis control fault during human-prosthesis interactive walking using intrinsic sensors. *IEEE robotics and automation letters*, 7(3):8307–8314, 2022.

[90] Benjamin A Newman, Reuben M Aronson, Siddhartha S Srinivasa, Kris Kitani, and Henny Admoni. Harmonic: A multimodal dataset of assistive human–robot collaboration. *The International Journal of Robotics Research*, 41(1):3–11, 2022.

[91] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *sensors*, 12(2):1211–1279, 2012.

[92] Scott Niekum, Sarah Osentoski, Christopher G Atkeson, and Andrew G Barto. Online bayesian changepoint detection for articulated motion models. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 1468–1475. IEEE, 2015.

[93] Ini Oguntola, Joseph Campbell, Simon Stepputtis, and Katia Sycara. Theory of mind as intrinsic motivation for multi-agent reinforcement learning. *arXiv preprint arXiv:2307.01158*, 2023.

[94] Allison M Okamura. Methods for haptic feedback in teleoperated robot-assisted surgery. *Industrial Robot: An International Journal*, 31(6):499–508, 2004.

[95] OpenAI. Swarm: Lightweight orchestration for multi-agent systems. `https://github.com/openai/swarm`, 2024. Accessed: May 17, 2025.

[96] Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903. IEEE, 2024.

[97] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019.

[98] Friedemann Pulvermüller. Brain mechanisms linking language and action. *Nature reviews neuroscience*, 6(7):576–582, 2005.

[99] David Quintero, Anne E Martin, and Robert D Gregg. Toward unified control of a powered prosthetic leg: A simulation study. *IEEE Transactions on Control Systems Technology*, 26(1):305–312, 2017.

[100] Alec Radford et al. Clip: Connecting text and images. In *ICML*, 2021.

[101] Muhammad A Rahman, Niklas Hopner, Filippos Christianos, and Stefano V Albrecht. Towards open ad hoc teamwork using graph-based policy learning. In *International conference on machine learning*, pages 8776–8786. PMLR, 2021.

[102] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

[103] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, et al. A generalist agent. *arXiv preprint arXiv:2205.06175*, 2022.

[104] Jacob Rosen and Joel C Perry. Upper limb powered exoskeleton. *International Journal of Humanoid Robotics*, 4(03):529–548, 2007.

[105] Jonas Rothfuss, Dominique Heyn, Andreas Krause, et al. Meta-learning reliable priors in the function space. *Advances in Neural Information Processing Systems*, 34:280–293, 2021.

[106] Baltej Singh Rupal, Sajid Rafique, Ashish Singla, Ekta Singla, Magnus Isaksson, and Gurvinder Singh Virk. Lower-limb exoskeletons: Research trends and regulatory guidelines in medical and non-medical applications. *International Journal of Advanced Robotic Systems*, 14(6):1729881417743554, 2017.

[107] Tim Salzmann, Elia Kaufmann, Jon Arrizabalaga, Marco Pavone, Davide Scaramuzza, and Markus Ryll. Real-time neural mpc: Deep learning model predictive control for quadrotors and agile robotic platforms. *IEEE Robotics and Automation Letters*, 8(4):2397–2404, 2023.

[108] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.

[109] Robin Schmid, Deegan Atha, Frederik Schöller, Sharmita Dey, Seyed Fakoorian, Kyohei Otsu, Barry Ridge, Marko Bjelonic, Lorenz Wellhausen, Marco Hutter, et al. Self-supervised traversability prediction by learning to reconstruct safe terrain. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12419–12425. IEEE, 2022.

[110] Aimee E Schultz and Todd A Kuiken. Neural interfaces for control of upper limb prostheses: the state of the art and future possibilities. *PM&R*, 3(1):55–67, 2011.

[111] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. Joint action: bodies and minds moving together. *Trends in cognitive sciences*, 10(2):70–76, 2006.

[112] Sheng Shen, Tianqing Zhu, Di Wu, Wei Wang, and Wanlei Zhou. From distributed machine learning to federated learning: In the view of data privacy and security. *Concurrency and Computation: Practice and Experience*, 34(16):e6002, 2022.

[113] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[114] Madhuri Singh, Amal Alabdulkarim, Gennie Mansi, and Mark O Riedl. Explainable reinforcement learning agents using world models. *arXiv preprint arXiv:2505.08073*, 2025.

[115] SS Srinivasan, MJ Carty, PW Calvaresi, TR Clites, BE Maimon, CR Taylor, AN Zorzos, and H Herr. On prosthetic control: A regenerative agonist-antagonist myoneural interface. *Science Robotics*, 2(6):eaan2971, 2017.

[116] Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1504–1509, 2010.

[117] Jakob Suchan and Jan-Patrick Osterloh. Assessing drivers' situation awareness in semi-autonomous vehicles: Asp based characterisations of driving dynamics for modelling scene interpretation and projection. *arXiv preprint arXiv:2308.15895*, 2023.

[118] György Szarvas, Richárd Farkas, and Róbert Busa-Fekete. State-of-the-art anonymization of medical records using an iterative machine learning framework. *Journal of the American Medical Informatics Association*, 14(5):574–580, 2007.

[119] Durk Talsma. Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in integrative neuroscience*, 9:19, 2015.

[120] Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[121] Michael Tomasello and Hannes Rakoczy. What makes human cognition unique? from individual to shared to collective intentionality. *Mind & language*, 18(2):121–147, 2003.

[122] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[123] Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*, 2025.

[124] Michael R Tucker, Jeremy Olivier, Anna Pagel, Hannes Bleuler, Mohamed Bouri, Olivier Lambercy, José del R Millán, Robert Riener, Heike Vallery, and Roger Gassert. Control strategies for active lower extremity prosthetics and orthotics: a review. *Journal of neuroengineering and rehabilitation*, 12:1–30, 2015.

[125] Mudit Verma, Siddhant Bhambri, and Subbarao Kambhampati. Theory of mind abilities of large language models in human-robot interaction: An illusion? In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 36–45, 2024.

[126] Jörn Vogel, Annette Hagengruber, Maged Iskandar, Gabriel Quere, Ulrike Leipscher, Samuel Bustamante, Alexander Dietrich, Hannes Höppner, Daniel Leidner, and Alin Albu-Schäffer. Edan: An emg-controlled daily assistant to help people with physical disabilities. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4183–4190. IEEE, 2020.

[127] Binglu Wang, Lei Zhang, Zhaozhong Wang, Yongqiang Zhao, and Tianfei Zhou. Core: Cooperative reconstruction for multi-agent perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8710–8720, 2023.

[128] Rujia Wang, Xiangbo Gao, Hao Xiang, Runsheng Xu, and Zhengzhong Tu. Cocmt: Communication-efficient cross-modal transformer for collaborative perception. *arXiv preprint arXiv:2503.13504*, 2025.

[129] Yue Wen, Jennie Si, Andrea Brandt, Xiang Gao, and He Helen Huang. Online reinforcement learning control for the personalization of a robotic knee prosthesis. *IEEE transactions on cybernetics*, 50(6):2346–2356, 2019.

[130] Jonathan R Wolpaw. Brain–computer interfaces. In *Handbook of clinical neurology*, volume 110, pages 67–74. Elsevier, 2013.

[131] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

[132] Daniel M Wolpert. Computational approaches to motor control. *Trends in cognitive sciences*, 1(6):209–216, 1997.

[133] Daniel M Wolpert, Zoubin Ghahramani, and Michael I Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, 1995.

[134] Daniel M Wolpert and Mitsuo Kawato. Multiple paired forward and inverse models for motor control. *Neural networks*, 11(7-8):1317–1329, 1998.

[135] Michael Wooldridge. *An introduction to multiagent systems*. John wiley & sons, 2009.

[136] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 3(4), 2023.

[137] Jiaxu Xing, Ismail Geles, Yunlong Song, Elie Aljalbout, and Davide Scaramuzza. Multi-task reinforcement learning for quadrotors. *IEEE Robotics and Automation Letters*, 2024.

[138] Mengwei Xu, Feng Qian, Qiaozhu Mei, Kang Huang, and Xuanzhe Liu. Deeptype: On-device deep learning for input personalization service with minimal privacy concern. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–26, 2018.

[139] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022.

[140] Xingyu Yang, Yixiong Du, Leihui Li, Zhengxue Zhou, and Xuping Zhang. Physics-informed neural network for model prediction and dynamics parameter identification of collaborative robot joints. *IEEE Robotics and Automation Letters*, 8(12):8462–8469, 2023.

[141] Xingyu Yang, Zhengxue Zhou, Leihui Li, and Xuping Zhang. Collaborative robot dynamics with physical human–robot interaction and parameter identification with pinn. *Mechanism and Machine Theory*, 189:105439, 2023.

[142] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, et al. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21361–21370, 2022.

[143] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, 2021.

[144] Yichen Zhu, Zhicai Ou, Xiaofeng Mou, and Jian Tang. Retrieval-augmented embodied agents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17985–17995, 2024.