Evolution of Humanoid Locomotion Control

Yan Gu[1,⋆], Guanya Shi[2,⋆], Fan Shi[3,⋆], I-Chia Chang[1], Yen-Jen Wang[4], Qilong Cheng[5], Zachary Olkin[6], Ivan Lopez-Sanchez[5], Yunchu Feng[5], Jian Zhang[7], Aaron D. Ames[6], Hao Su[5,†], and Koushil Sreenath[4,†]

[1]School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA.

[2]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

[3]Department of Electrical and Computer Engineering, National University of Singapore, Singapore.

[4]Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA 94720 USA.

[5]Tandon School of Engineering, New York University, New York, NY 11201 USA.

[6]Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

[7]Meta Platforms Inc., Menlo Park, CA 94025, USA.

[†]Corresponding authors. Emails: hao.su@nyu.edu, koushils@berkeley.edu

[⋆]These authors contributed equally to this work.

Humanoid robots stand at the forefront of robotics, aiming to capture the agility, robustness, and expressivity of human movement within an anthropomorphic form. The locomotion control of humanoids has evolved from classical model-based methods to reinforcement learning powered by large-scale simulation, and now to generative models that produce adaptive, whole-body behaviors, propelling humanoids toward operation in real-world environments. This survey positions humanoid control at a turning point, converging toward a unified paradigm of physics-guided generative intelligence that integrates optimization, learning, and predictive reasoning. We identify three core principles linking these paradigms: physics-based modeling, constrained decision making, and adaptation to uncertainty. Building on these connections, we provide recommendations for researchers and outline open challenges in safety, accessibility, and human-

level capability. These directions represent a transformation from engineered stability to intelligent autonomy, laying the groundwork for humanoid generalists capable of operating safely, collaborating naturally, and extending human capability in the open world.

Summary

Humanoid locomotion control is evolving from physics-based feedback toward generative, adaptive, and human-compatible intelligence.

Introduction

Humanoid robots represent one of the most ambitious frontiers in robotics, integrating human-like agility, robustness, and expressivity within an anthropomorphic body. Equipped with legs and arms, they can climb stairs, traverse cluttered environments, and collaborate safely with people. Unlike wheeled machines, humanoids can step over obstacles, adapt to discontinuous or moving terrain, and operate in spaces designed for humans. Their human-like morphology narrows the embodiment gap and enables direct use of human motion data and control strategies, making humanoids a natural platform for connecting high-level reasoning with whole-body motor execution in existing human environments.

The long-term vision is to realize humanoids that coexist safely and productively with people, robots that are physically capable, perceptually aware, and socially intelligible. Achieving this vision, however, remains a grand challenge. Coordinating dozens of joints and intermittent contacts under uncertain, nonlinear dynamics requires controllers that prioritize safety while balancing precision, energy efficiency, and adaptability in unpredictable environments.

A central challenge to realizing this vision is control, the mechanism that makes hardware produce purposeful behavior. Control for bipedal humanoid locomotion has advanced rapidly over the past several decades (**Fig. 1**). Early approaches relied on simplified physics-based dynamics models and handcrafted feedback laws to ensure stability. Advances in computation enabled nonlinear controller synthesis through offline trajectory optimization using detailed dynamic models, while predictive control introduced constraint handling for online motion generation of robust whole-body behaviors. Recently, Reinforcement Learning (RL) has demonstrated highly agile and resilient locomotion through large-scale simulation, parallel training, and physics-guided reward design. Emerging

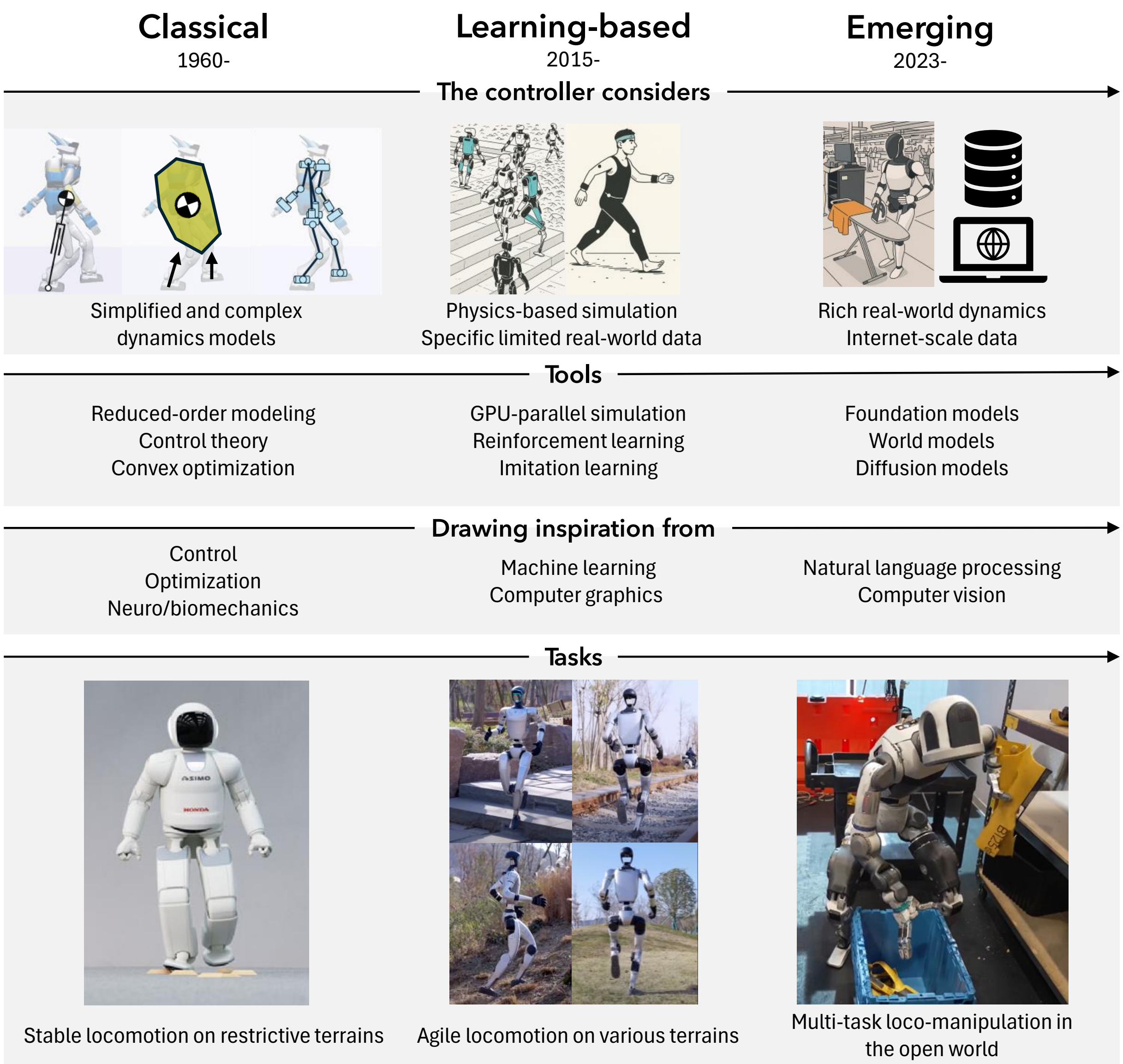

Figure 1: Evolution of humanoid locomotion control paradigms over the past several decades. Humanoid control has progressed from classical methods, including real-time feedback and predictive control based on simplified dynamics, control theory, and convex optimization, to learning-based approaches that leverage physics simulation, parallel computation, and reinforcement learning. It is now entering an emerging era driven by generative AI. Each paradigm expands the scope of reasoning from reduced-order models to simulated and real-world physics and increasingly integrates insights from multiple disciplines, including biomechanics, control theory, machine learning, computer graphics, natural language processing, and computer vision. This progression has enabled a transition from stable walking on constrained terrains to agile locomotion and, most recently, to multi-task loco-manipulation in unstructured environments.

generative methods now unify perception, high-level decision-making, and control to enable versatile, adaptive, and human-like motion.

This survey examines the evolution of humanoid locomotion control paradigms from classical control to RL and generative models, driven by the need for greater robustness, adaptability, and generalization in complex, real-world environments (**Fig. 1**, **Tab. 1**). It reveals that progress across these paradigms is unified by enduring physics-based principles that continue to guide the development of intelligent, physically adept humanoids. We also identify open challenges in safety, accessibility, and cognitive capability essential for deployment beyond the laboratory. Perception (*1, 2*), task planning (*3*), and state estimation (*4*) are discussed only when directly relevant to control design.

Unlike prior surveys that focus on individual control paradigms or specific application domains (*5, 6, 7, 8, 9, 10, 11, 12, 13*), this paper bridges classical and learning-based approaches through their common physical principles and extends the discussion to recent advances in generative and foundation models. Rather than serving solely as a retrospective review, it offers a forward-looking perspective on how these paradigms are converging toward unified principles of humanoid control. Ha et al. (*13*) focus on RL for quadrupeds, Wensing et al. (*14*) review optimization-based control without addressing learning approaches, and Gu et al. (*15*) provide an overview of whole-body loco-manipulation emphasizing system integration rather than the conceptual evolution of control frameworks.

Contributions and key takeaways. This paper frames the advancements of humanoid locomotion control from physics-based to learning-based methods with a view toward their fusion within emerging paradigms for next-generation humanoids. Historically, physics-based control provided stability, interpretability, and guarantees, while learning-based methods have delivered agility and robustness. Our first contribution unifies these perspectives (**Fig. 1**) by identifying their respective strengths and limitations: classical real-time feedback control is fast and principled but restricted to pre-planned behaviors; predictive control manages constraints online but often sacrifices agility; RL achieves agility and robustness but remains task-specific and opaque; and generative methods offer versatility yet lack real-world reliability.

The second contribution reveals the inherent connections among these paradigms through three shared enablers: physics-based modeling, constrained decision-making, and adaptation to uncertainty. Each method excels in a particular domain, and their integration promises new capabilities for humanoid systems. We posit that the field is converging toward controllers that unify classical approaches with learning, predictive reasoning, and adaptation, where physics increasingly guides learning through reward design, system identification, and real-time interfaces. Lastly, we outline emerging directions in generative multi-modal loco-manipulation, test-time adaptation, and human-centered humanoid design. The future of humanoid locomotion control lies at this intersection, drawing from the rigor of the past, leveraging the developments of the present, and integrating the transformative methods now on the horizon.

Modeling and Classical Control for Humanoid Locomotion

This section reviews how classical control for bipedal humanoid locomotion has advanced from real-time feedback to predictive methods. This shift has been powered by the demand for improving robustness, agility, and explicit constraint handling in complex environments.

Early real-time feedback controllers operated myopically, reacting directly to measured states with minimal computation. Since the 1980s, such controllers have enabled remarkable physical demonstrations, including Raibert's dynamic hopping and jumping bipeds (*16*), Honda's quasi-static ASIMO (*17*), and later dynamic bipedal (*18, 19*) and humanoid (*20*) walking and running (*21*).

Predictive control extended these methods by reasoning about system dynamics and constraints over a finite horizon to plan motions proactively. Offline trajectory optimization, which generates desired full-body motions, enabled major milestones such as the DARPA Robotics Challenge (2013–2015) (*22, 23, 24*) and Boston Dynamics' ATLAS demonstrations of parkour behaviors (*25, 26*). Subsequent advances in online predictive control for quadrupeds (*27*) inspired optimization-based humanoid controllers (*14*), further enhancing agility and robustness.

Physics-based dynamics models underpin both classical and learning-based control, shaping achievable performance and computational requirements. Classical control, grounded in these models, has enabled agile and robust behaviors but struggles with generalization, state estimation, and scalability to unstructured environments. Learning-based methods have begun to overcome these challenges, yet their success often relies more on physics-guided training design than on algorithmic novelty, as discussed in "Learning-based Control for Humanoid Locomotion." Reward shaping and parameter randomization embed the same physics-informed priors and constraint reasoning that characterize classical control (*28, 29*). Ultimately, continued progress in both paradigms depends on accurate physics-based modeling, a central theme examined next.

Physics-based Analytical Models

Physics-based analytical models provide interpretable structure by describing how joint torques, contact forces, and environmental interactions generate motion, typically derived from first-principles dynamics. A general formalization for a single-domain hybrid dynamical system with state x and control input u has dynamics:

$$\Sigma := \begin{cases} \dot{x} = f(t, x, u) & x \in D \setminus S & \text{(1a)} \\ x^{+} = \Delta(x^{-}) & x^{-} \in S & \text{(1b)} \end{cases}$$

Here t denotes time, and the continuous f and discrete Δ dynamics can be derived using Lagrangian or Newton-Euler formulations (*30*) given the domain D and switching surface S (extensions exist for multi-domain systems (*31*)). This hybrid structure is essential for legged systems, which repeatedly make and break ground contact. Any useful model (**Fig. 2**), analytical, numerical, or learned, must be accurate enough to reflect physical reality and efficient enough to run at control rates, which is difficult for high-dimensional, nonlinear, hybrid, and contact-rich humanoid dynamics.

Full-order models. Full-order models describe the complete rigid-body dynamics and environment interaction of the robot (*30*). The choice of contact defines holonomic constraints that shape the hybrid system's graph structure, making contact selection critical to modeling. While utilizing full-order models allows the controllers to achieve more dynamic motion, their high dimensionality increases computational cost and control complexity.

Reduced-order models. Reduced-order models compress behavior to a few task-relevant quantities such as center of mass (CoM) motion or centroidal momentum (*32*). Examples include the linear inverted pendulum (LIP) (*33*), spring-loaded inverted pendulum (SLIP) (*16, 34*)), compass-gait biped (*35*), centroidal models (*36*), point mass model (*37*), and single rigid body models (*27, 38*). These abstractions enable real-time balance control, foot placement, and disturbance rejection (*39, 40, 41, 42, 43, 44, 45*), while also revealing limb coordination principles in animal locomotion (*46*) and guided bio-inspired control strategies (*47*). Nonetheless, their simplifying assumptions, such as flat terrain, rigid contact, or constant CoM height (*33*), limit performance in fast, contact-rich motions and complex environments.

Physics-based simulators. Physics-based models (typically full-order) remain the invisible engine behind modern learning methods: RL owes much of its progress to versatile and fast physics simulators (*48, 49, 50, 51*), whose fairly accurate dynamics, ease of incorporating terrains and objects, and massive parallelization make training feasible at scale. Importantly, simulators implicitly handle discrete transitions allowing one to forgo explicit hybrid system models of locomotion.

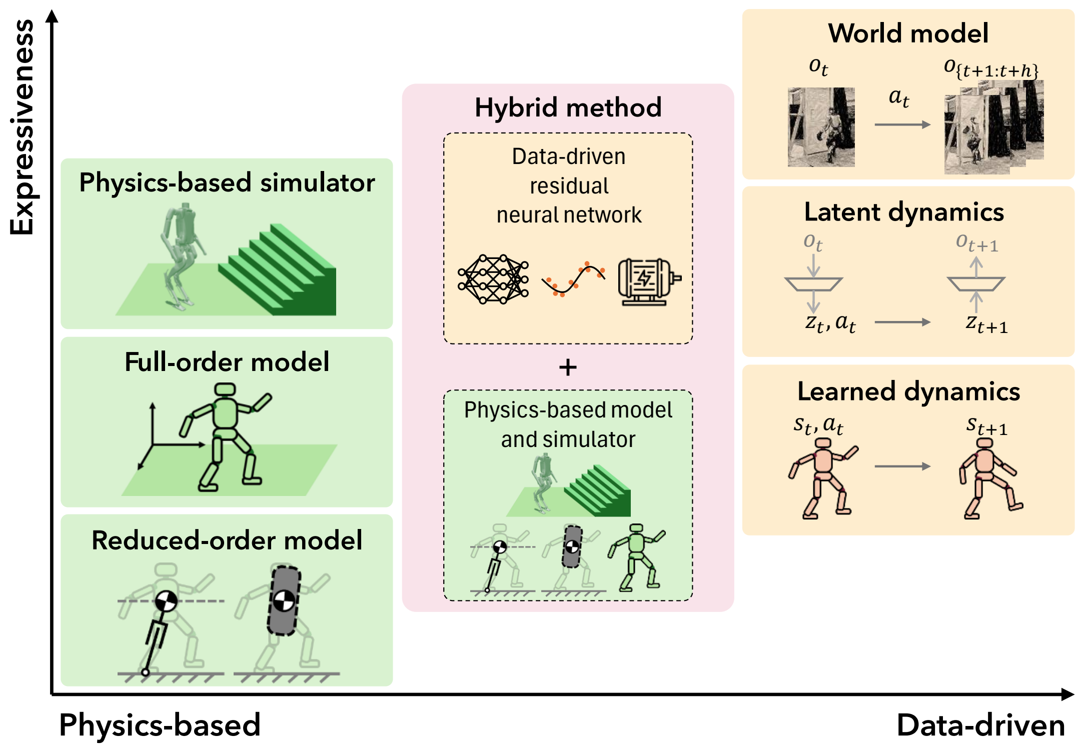

Figure 2: Spectrum of modeling approaches for humanoid locomotion control, spanning from physics-based to data-driven methods and from low to high model expressiveness. Physics-based models include: reduced-order models (*33, 34, 38*), which approximate dynamics using simplifying assumptions; full-order models (*30*), which capture complete robot dynamics under specified contact conditions; and simulators (*48, 49*), which provide the highest fidelity through detailed modeling of contact, actuation, and sensing (reviewed in "Modeling and Classical Control for Humanoid Locomotion"). Data-driven models comprise: learned dynamics (*52, 53*), which model state transitions directly from data; latent dynamics models (*54*), which operate in lower-dimensional space; and world models (*55, 56, 57*), which predict high-dimensional robot and environment behavior (see "Learning-based Control for Humanoid Locomotion" and "Emerging Frontiers in Humanoid Robot Control"). Hybrid approaches integrate physics-based and learned components to exploit complementary strengths, such as modeling actuator dynamics (*58*), learning residual effects (*59, 60*), or improving simulator fidelity with real-world data (*61*). Color shading indicates different modeling approaches: classical , learning-based , hybrid , and emerging . Symbols o_t, s_t, z_t, and a_t represent the observation, state, latent state, and action at time t, respectively, while $o_{\{t+1:t+h\}}$ denotes a prediction of observations from $t+1$ to $t+h$ where h is the time horizon.

Classical Real-time Feedback Control of Humanoid Locomotion

The real-time control layer for legged robots operates at high loop rates and without a prediction horizon. Early successes in bipedal locomotion were achieved through such real-time feedback control, which stabilized motion using physics-based models and handcrafted feedback laws. These classical controllers enabled the first demonstrations of stable walking and balance recovery but were limited in robustness and adaptability. Still, these controllers remains indispensable today, serving as the low-level layer that tracks motion commands in all control architectures, including learning-based (see "Learning-based Control for Humanoid Locomotion") and emerging approaches (see "Emerging Frontiers in Humanoid Robot Control"), as illustrated in **Fig. 3**. Classical feedback methods are broadly categorized into linear approaches for reduced-order models and nonlinear approaches for full-order models; see (*5*) for a comprehensive review of real-time locomotion control.

Linear feedback control. Early biped control replied on simplifying assumptions to derive analytical models and controllers. The Zero Moment Point (ZMP) method ((*62, 63*)) generated CoM trajectories to ensure stable quasi-static walking for fully actuated humanoids, often based on the LIP model (*63*). Raibert's control method (*16*) enabled dynamic hopping and running behaviors for bipeds based on compliant reduced-order models and foot placement. The capture point approach (*64*) bridged CoM and foot placement control by identifying where to step to stop motion, typically using LIP for efficiency. Subsequent methods generalized feedback control on reduced-order models using optimal control such as linear quadratic regulators (*65*). Lastly, the simplest linear method is the Proportional-Derivative (PD) control, often combined with inverse kinematics or dynamics for classical motion tracking (*40, 66, 67, 68*). In learning-based systems, PD target tracking (*69*) serves as the most popular interface between RL and real-time control: the RL policy outputs desired joint states, and a PD controller tracks them in real time (**Fig. 3B**). Future work may replace PD control with more advanced controllers for improved performance.

Nonlinear feedback control. The simplifying assumptions of reduced-order models limit controller applicability. Full-order models remove these assumptions but require handling hybrid, nonlinear robot dynamics (*30, 70*). This is typically achieved by decomposing the system into actuated and passive (zero) dynamics and applying nonlinear control (*30*). Nonlinear full-order models can also be addressed by optimization-based control, such as embedding control Lyapunov functions (CLF) in a Quadratic Program (QP) (i.e., CLF-QP (*71, 72*)) and whole-body QPs (*22, 67, 73*). They can track desired motions in real time but remain "greedy" in that they do not use a prediction horizon.

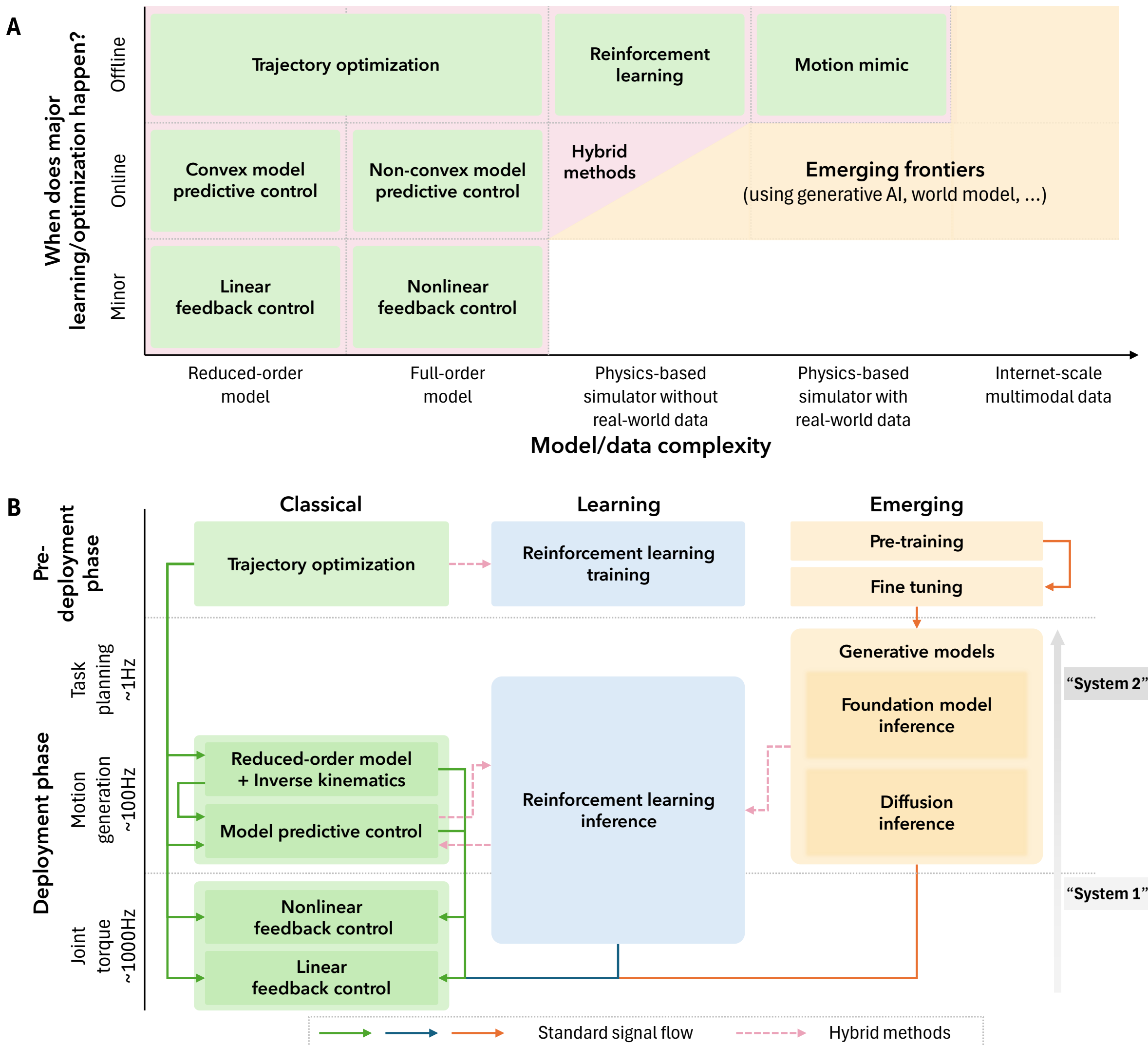

Figure 3: Unified view of humanoid locomotion control paradigms across model–data complexity and computational hierarchy. (**A**) Categorization of control approaches by when optimization or learning occurs (minor, online, or offline) and by model or data complexity. (**B**) Organization of control paradigms by their role in pre-deployment versus deployment phase, with corresponding control loop rates. Both views reveal a consistent progression from classical to learning-based and emerging paradigms. Classical real-time feedback controllers, including linear and nonlinear methods, involve minimal optimization or learning and operate at high frequencies ($\sim$1000 Hz) but offer limited foresight and adaptability. Predictive control methods, such as model predictive control, reason about future states online ($\sim$100 Hz) using physics-based models. Learning-based controllers, such as RL, are trained offline in simulation and deployed onboard at comparable rates. Emerging generative approaches leverage large-scale data and hybrid training, typically operating at lower deployment frequencies due to computational demands. The diagram also highlights two levels of reasoning: "System 1," governing fast, reactive motor execution, and "System 2," capturing slower, deliberative reasoning. This organization suggests a future integration of rapid feedback and high-level decision-making in humanoid control. Color shading indicates control paradigms: classical, learning-based, hybrid, and emerging.

Predictive Control for Humanoid Locomotion

Advances in computation have enabled predictive control, a paradigm that optimizes behavior over a time horizon to enhance robustness beyond instantaneous feedback. These methods are generally formulated as optimization problems that minimize a cost function $L(t, x, u)$, composed of a running cost l and terminal cost V at the terminal time T, subject to system dynamics (discussed in "Physics-based Analytical Models") and hard constraints, solved either offline through trajectory optimization or online via Model Predictive Control (MPC). The optimization yielding control inputs u and states x is expressed as:

$$\begin{aligned} \underset{u,x}{\text{minimize}} \quad & L(t,x,u) = \int l(t,x,u)dt + V(x(T)) && \text{(Objective)} \quad \text{(2a)} \\ \text{subject to} \quad & \dot{x} = f(t,x,u) && \text{(Dynamics)} \quad \text{(2b)} \\ & x \in \mathcal{X},\ u \in \mathcal{U} && \text{(Constraints)} \quad \text{(2c)} \end{aligned}$$

Here $\mathcal{X}$ and $\mathcal{U}$ denote the admissible state and input sets, encompassing joint limits, contact constraints associated with the discrete model component (1b), and actuator bounds. For a comprehensive review of optimization-based methods, see (*14*).

Offline trajectory optimization. Trajectory optimization was a common tool for long-horizon planning, that is, planning over extended time horizons or across multiple footsteps, used prominently in the DARPA Robotics Challenges (*23*). One successful strategy during the event involved two stages: (1) determining footstep locations and (2) planning continuous motions consistent with those steps (*22, 24, 26*). Powerful solvers such as differential dynamic programming, sequential QP, and mixed-integer convex optimization were employed to plan both footstep locations and motions based on reduced-order models (*22, 26*), while later approaches unified them into a single optimization (*74*). Subsequent work has incorporated upper-limb coordination, allowing hand contacts to enhance whole-body stability on challenging terrain (*75*). Hybrid Zero Dynamics (HZD) is another trajectory optimization approach that stabilizes dynamic bipedal locomotion by explicitly addressing the robot's full-order, nonlinear, hybrid, and underactuated dynamics (*18, 20, 30, 31, 70, 76*). HZD can also generate human-like gait by embedding human movement patterns into the optimization (*76*), mirroring the use of human data in RL as reviewed in

"Learning from Real-World Human and Robot Data".

Online model predictive control. Solving trajectory optimization problems online introduces strict computational and timing constraints, which tends to limit horizons to under one second (*77*). To reduce the dimensionality of the problem, controllers often use reduced-order models such as single rigid body or centroidal approximations (*78, 79*). When nonlinear dynamics are considered, solving only one QP within a sequential QP framework enables real-time execution (*77, 80*).

To enhance robustness, MPC can adapt foot-contact locations and timings online rather than only optimizing continuous actions. One example is contact-implicit MPC (*81, 82, 83, 84*), which jointly optimizes contacts and motions via gradient-based solvers. Alternatively, sampling-based MPC employs gradient-free optimization to determine system inputs and implicitly choose contacts (*85, 86, 87*). Yet, most humanoid implementations are limited to simulation, and achieving computational tractability in real-world deployment remains an open challenge.

Summary of Performance Boundaries of Classical Control

Classical control methods face fundamental limitations in real-world deployment. First, they rely on accurate state estimation and lack robustness under sensor noise, partial observability, or perceptual delays. Second, a persistent modeling gap also constrains performance: reduced-order models are fast but miss key limb and contact dynamics during agile motion, while full-order models capture these effects but are sensitive to modeling errors and environmental uncertainty such as terrain compliance and friction variation. Third, the core trade-off between model accuracy and computational speed remains unresolved: controllers based on detailed models are often too slow for onboard deployment, while simplified models may prevent agility. Finally, classical approaches require extensive manual tuning and strong contact assumptions, limiting adaptability to new tasks or environments. These challenges motivate learning-based approaches that can handle uncertainty and partial observability, generalize across conditions, and use data to close the gap between model assumptions and real-world complexity, while still drawing on principles from classical control as discussed next.

Table 1: Summary of humanoid locomotion control paradigms and their core principles.

Category	*Main methods*	*Core ideas*	*Year*
Classical			
Linear feedback control	ZMP (*62, 63*); Raibert's control (*16*); capture point (*64*)	Using reduced-order models and employing linear systems theory to regulate locomotion behavior in real time.	1980s–
Nonlinear feedback control	Inverse dynamics (*66*); CLF-QP (*71, 72*); whole-body QP (*22, 67, 73, 88*)	Using full-order models and applying nonlinear control theory to produce joint torques for whole-body motion tracking.	2000s–
Trajectory optimization	HZD (*30*); methods using reduced-order models (*22, 23, 24, 26, 89*)	Long-horizon planning that selects footsteps and optimizes CoM and whole-body trajectories based on analytical models.	2010s–
Model predictive control	Methods using reduced-order models (*78*); contact-implicit MPC (*81*); sampling-based (*85*)	Short-horizon predictive control with online replanning; reduced-order methods for real-time control; contact-implicit MPC and sampling-based methods for contact selection.	2010s–
Learning-based			
Reinforcement learning with pure simulation	Policy gradient (*90, 91*); curriculum learning (*92*); teacher-student (*93, 94*)	Utilizing physics-based simulation and massive-parallel training to learn a policy offline and generate motion plans online.	2016–
Reinforcement learning with real-world data	Motion imitation (*95, 96, 97*); residual learning (*61*)	Motion imitation: using human motion data as reference to reduce the need for manually designed reward functions. Residual learning: mitigating the sim-to-real gap caused by modeling errors by refining both the simulator and the learned policy using real-world data.	2018–
Model-guided reinforcement learning	Tracking trajectories (*98, 99, 100*); controllers for reward shaping (*101, 102*)	Using model-based methods to guide the learning and reduce dependence on heuristics; embedding trajectories or controllers into the RL reward.	2020–
Emerging			
LLM / VLM guidance	LLM / VLM-conditioned control (*103, 104, 105*)	Conditioning control on language and vision for goal grounding and semantic reasoning for high-level intent planning.	2024–
Diffusion motion models	Generative gait and trajectory priors (*97, 106*)	Diffusion priors sample diverse and feasible humanoid motions used as tracking and regularization targets.	2025–
World / foundation models	Latent predictive dynamics (*107, 108*)	Learning latent dynamics from data that predicts future observations, rewards, and contact events, enabling control via imagined rollouts in latent space.	2025–

Learning-based Control for Humanoid Locomotion

Conventional RL algorithms, developed since the 1980s (*109, 110, 111, 112, 113*), struggled to scale in high-dimensional continuous spaces due to limited function approximation and costly data collection (*114, 115, 116, 117*). Neural networks (NN) have been integrated into RL since around the 1990s (e.g., TD-Gammon (*118*) and early NN-based RL (*119, 120*)), but the 2013-2015 rise of deep learning (*121, 122*) triggered modern Deep RL (DRL). Breakthroughs such as Deep Q-Networks (*123, 124*), AlphaGo (*125*), and continuous control in MuJoCo (*90*) marked the shift toward scalable policy learning.

Translating DRL to humanoid locomotion remained challenging due to high-dimensional continuous actions, contact-rich dynamics, and costly failure. Key obstacles include slow or inaccurate simulators, expensive and risky real-world data collection, and insufficient compute for large-scale training. Recent advances have mitigated these barriers. Modern GPU simulators such as NVIDIA Isaac Sim (*49*) provide fairly realistic dynamics and thousands of parallel environments, cutting training time by orders of magnitude. Combined with open-source DRL libraries (*126, 127, 128*), robust hardware, and improved algorithms, learning-based control now exceeds classical approaches in locomotion agility and robustness (**Fig. 1**). Notably, RL for quadrupeds (*58, 129, 130*) and bipeds (*131, 132, 133, 134*) paved the way for humanoids, addressing the challenge of sim-to-real transfer. Subsequent work incorporated human motion data during training (*95, 97, 135, 136, 137*), improving realism and efficiency. These advances have enabled extreme behaviors such as parkour, human motion imitation, and dynamic loco-manipulation (e.g., humanoid table tennis (*138, 139*)).

This section reviews learning-based control for humanoid locomotion, focusing on simulation-driven RL, learning from real-world data, and hybrid approaches integrating RL with model-based control (**Fig. 3**). It also highlights the architectural and conceptual connections between learning-based and classical control paradigms (**Tab. 2**).

Learning from Pure Simulation

RL enables humanoid locomotion controllers to learn agile and robust behaviors directly from interaction data, and simulation provides a safe, efficient environment for exploring diverse terrains and disturbances before real-world deployment. RL-based humanoid locomotion control can be formulated as a Partially Observable Markov Decision Process (POMDP) (*140*). A POMDP is defined as $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, p, r, \mathcal{O}, \gamma \rangle$, where $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{O}$ are the admissible state, action, and observation sets. Here $p(s'|s, a)$ denotes the state transition function capturing the system dynamics that governs the evolution from state s and action a to state s', $r(s, a)$ is the reward function, and γ is the discount factor. The agent learns a policy $\pi(a_t|o_t)$ that maps observations o_t to action a_t at timestep t to maximize the expectation of the discounted cumulative reward $J(\pi)$, which encourages long-horizon performance under stochastic dynamics:

$$
\begin{aligned}
&\underset{\pi}{\text{maximize}} \quad && J(\pi) = \mathbb{E}_{\tau \sim p_\tau(\cdot|\pi)} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] && \text{(Objective)} && \text{(3a)} \\
&\text{subject to} \quad && s_{t+1} \sim p(\cdot \mid s_t, a_t), && \text{(Dynamics)} && \text{(3b)} \\
& && a_t \sim \pi(\cdot \mid o_t), && \text{(Action model)} && \text{(3c)} \\
& && o_t \sim \mathcal{O}(\cdot \mid s_t), && \text{(Observation model)} && \text{(3d)} \\
& && s_t \in \mathcal{S},\ a_t \in \mathcal{A},\ o_t \in \mathcal{O}, && \text{(Constraints)} && \text{(3e)}
\end{aligned}
$$

where $\tau = (s_0, o_0, a_0, r_0, s_1, \dots)$ denotes a trajectory sampled from the trajectory distribution $p_\tau(\cdot \mid \pi)$ induced by policy π and the environment dynamics $p(s'|s, a)$. The observation model $\mathcal{O}(o_t|s_t)$ reflects partial observability. For an in-depth discussion of POMDP and RL, see (*13, 113*).

To enable efficient training and real-world deployment, recent research has introduced a suite of complementary techniques. The rest of this subsection reviews advances in parallelized simulation platforms, open-source learning frameworks, efficient policy training algorithms, and sim-to-real transfer strategies, which collectively enhance scalability, reproducibility, and robustness in RL-based humanoid locomotion control.

Parallelized GPU-accelerated simulators. RL typically requires extensive trial-and-error interactions, making data collection a major training bottleneck. To address this, the field has seen a significant shift in simulation frameworks, from traditional CPU-based engines, such as RaiSim (*50*) and Pybullet (*51*), to massively parallel GPU-accelerated platforms, including NVIDIA Isaac series (*49*) and MuJoCo MJX (*141*). These modern simulators can execute thousands of environment rollouts simultaneously on a single consumer-grade GPU, improving simulation throughput by several orders of magnitude and making rapid experimentation far more accessible.

Policy gradient methods and open-source frameworks. Advances in policy gradient algorithms, such as Trust Region Policy Optimization (*90*) and Proximal Policy Optimization (*91*), have enabled stable and sample-efficient learning in legged locomotion, building on successes like AlphaGo (*125*). Together with high-quality open-source libraries (*128, 141, 142, 143*), these advances have made RL research more accessible, reproducible, and customizable. Shared pipelines and pretrained models reduce overhead for newcomers, accelerate experimentation, and promote fair comparisons across methods.

Domain randomization. Domain randomization (*144*) improves sim-to-real transfer by training policies under randomized conditions such as sensor noise, actuator latency, external forces, and model parameters. By exposing the policy to broad variations in simulation, it becomes more robust to real-world uncertainties. Carefully designed randomization schemes have demonstrated significantly improved transfer success in legged locomotion (*129,145*).

Curriculum learning. Learning locomotion in high-dimensional, dynamic environments remains challenging. Curriculum learning (*146*) addresses this by gradually increasing task complexity, allowing policies to acquire basic motor skills before adapting to complex scenarios. For instance, training can begin on flat terrain without disturbances and progress to uneven terrain or external pushes. Adaptive curriculum design further adjusts difficulty based on policy performance during training (*92, 134, 147*).

Teacher-student and asymmetric actor-critic frameworks. The teacher-student framework (*93,148*) improves learning by using privileged information, such as terrain friction, contact states, or external forces available only during training. The teacher policy, with full information, learns robust behaviors, while a student policy restricted to onboard sensors is trained to imitate the teacher. To improve generalization, a reconstruction loss can help the student infer unobservable features, accelerating training and benefiting sim-to-real transfer (*94, 147*). Similarly, the asymmetric actor-critic framework (*149*) assigns privileged information to the critic while limiting the actor (policy) to onboard observations. This asymmetry has recently gained attention for its simplicity and strong performance (*150, 151*).

Learning from Real-World Human and Robot Data

Learning from real-world human data. Because humanoid locomotion involves high-dimensional actions and complex limb coordination, simple reward terms including velocity tracking or gait frequency often yield unnatural behaviors, such as straight knees, foot scuffing, or asymmetric gait (*135, 152*). To address this, researchers increasingly utilize human motion priors to guide humanoid skill learning.

In computer graphics and animation, a variety of public motion capture datasets (*153, 154, 155, 156, 157, 158, 159, 160*) provide valuable priors for this purpose. Due to the embodiment gaps between humans and humanoids, a retargeting step (*136, 161, 162, 163, 164*) is typically required to convert human motion into kinematically feasible references. Once motion references are obtained, they can be incorporated into RL through two main strategies. Indirect (adversarial) methods, such as GAIL (*165*) and AMP (*135, 166*), learn discriminators to align robot motion distributions with those in reference data. Direct tracking methods, such as DeepMimic and its variants (*61, 95, 97*), H2O (*96, 136*), HumanPlus (*167*), ExBody (*168*), and general tracking methods (*169, 170, 171, 172*), define rewards explicitly as errors between reference and robot motions. Tracking-based schemes also extend to loco-manipulation, where RL agents jointly track robot, object, and contact references (*100, 163, 173, 174*).

System identification and actuator model learning from real-world data. While physics-based simulation reduces the gap between simulation and hardware, some parameters such as inertia are difficult to identify, and actuators or transmission dynamics remain hard to model due to nonlinearities and mechanical imperfections. Real-world data is therefore essential for improving simulation fidelity.

Because locomotion exhibits non-smooth hybrid dynamics, sampling-based identification methods are often preferred (*131, 175*). For actuators, NN models trained on hardware data can predict actual torque output, capturing actuator behavior more accurately and enhancing simulation realism (**Fig. 2**). Studies on the ANYmal quadruped show that such hybrid modeling improves simulation accuracy and control performance (*58, 128, 176*).

Learning robot dynamics from real-world data. System identification and actuator modeling remain limited by predefined structures. System identification requires prior knowledge of which parameters to estimated, while actuator models focus only on local uncertainties. A broader family of approaches instead aims to learn full-body robot dynamics directly from real-world data (**Fig. 2**). Given the complexity of humanoid dynamics, residual learning is a common strategy, where a residual model is learned on top of nominal simulator dynamics models to account for unmodeled effects. For example, ASAP (*61*) and its variants (*177, 178*) learn residual action models that align simulation with real-world dynamics. Beyond residual learning, entire forward dynamics can also be learned from both simulated and real-world data to achieve safe navigation in complex environment (*52*).

Table 2: Core challenges in humanoid locomotion control and inherent connections across different paradigms. Emerging topics primarily refer to generative approaches such as diffusion policies, vision-language-action (VLA) models, and world models.

Key challenges of humanoid locomotion	***Key ideas to resolve the challenges shared by different approaches***		
	Classical	**Learning-based**	**Emerging**
Stabilizing unstable dynamics	Analytical dynamic modeling (*30, 33*); stabilizing controllers (*39, 63*)	Physics simulators (*48, 49*); learning-based controller (*128, 179*)	World models for dynamics, planning, and control (*54, 108, 180*)
High dimensionality	Reduced-order models (*32, 33*); hierarchical control frameworks (*40*)	Autoencoders, latent space (*56, 104*); latent dynamics (*181*)	Diffusion policies (*97, 182*); VLA models (*183, 184*)
Stabilizing hybrid dynamics	Hybrid system formulations (*30*); hybrid decision making (*81, 85*)	Learning hybrid automata (*185*)	Generative sequence models (*186, 187*); world models (*108*)
Constraints	Constrained optimization (*88*)	Reward shaping (*133, 188*)	Conditional diffusion (*97, 189*); VLA grounding (*105, 190*)
Robustness	Robust control (*191, 192*)	Domain randomization (*144, 145*)	Diffusion for diverse motion (*106*); large-scale VLA (*193*); world models (*194*)
Adaptation	Adaptive control (*44, 195*)	Adaptation modules (*196*)	In-context learning (*103*); diffusion test-time adaptation (*97*)

Inherent Connections to Classical Control

Building model-based systems requires expertise in dynamics, control, and optimization, whereas learning-based pipelines are often more accessible. However, classical principles remain essential, especially in RL components like domain randomization, reward design, and curriculum learning, all of which rely on understanding underlying dynamics. This section explores the connections between classical and learning-based control, offering a unified framework for humanoid locomotion (**Tab. 2**, **Fig. 3**).

An optimal control view point on RL-trained locomotion controller. As shown in (3) and **Fig. 3**, RL-trained locomotion controllers effectively solve a large number of optimization problems offline in simulation, and "encode" the solutions into NN policies. The optimization problems are solved across enormously diverse states and environments, producing policies that are adaptive, robust, and generalizable. Meanwhile, classical predictive control methods, formulated as in (2), often solve the problem *online* for a given state using simplified models (e.g., reduced-order model) to ensure fast computation and responsive behavior. In general, the resulting behavior is often sub-optimal due to model inaccuracies and the need for explicit state estimation. In contrast, RL policies can leverage large-scale GPU-accelerated offline training, sophisticated simulators, and frameworks such as teacher-student or asymmetric actor-critic learning to bypass explicit state estimation and improve generalization. However, as suggested by recent advances, leveraging both offline and online computation often leads to best results (details in “Emerging Frontiers in Humanoid Robot Control” and **Fig. 3B**).

Architecture and environment design in RL-based pipelines inspired by classical control. The connections between RL training and classical control are detailed next (**Tab. 2**). (i) Domain randomization (*69, 144*), which adds noise into the training process, aligns with principles of robust control (*191, 197*), where controllers are designed to ensure robustness across a range of system parameters. (ii) Adaptation module and teacher-student frameworks (*58, 196, 198*) reflect the idea of adaptive control (*44, 195, 199*), which estimates and compensates for uncertain system parameters online using historical state-action data. (iii) Finding the latent space, whether through modern methods such as autoencoders (*56, 181*) or classical techniques (e.g., principle component analysis (*200*), singular value decomposition (*201*), and using dynamics reduction through feedback (*40*)), follows the same goal of representing the high-dimensional robot dynamics using low-dimensional coordinates for efficient planning and control (*40, 56, 181*). (iv) The hybrid nature of locomotion, marked by discrete contact switching, has long been modeled as a hybrid dynamical system (*30*), inspiring structural design of learning policy, e.g., hybrid automata learning (*185, 202*), that explicitly handle multiple contact modes and transitions. (v) Hierarchical or cascaded control structures (*40*) common in classical approaches continue to inform hierarchical learning frameworks (*138*), offering interpretable interfaces, modularity and reuse, and improved sim-to-real transfer.

Merging classical control and RL for performance, robustness, and safety. Tuning RL policies for humanoid locomotion is often time-consuming due to the need for carefully designed rewards, environments, and curricula. To address this challenge, model-guided approaches leverage physics-based or control-theoretic insights to reduce heuristics while improving precision, robustness, and safety. Reduced-order models (e.g., SLIP (*98*) and single rigid body (*38*)) and full-order models (*99, 100*) have been used to generate reference motions for RL tracking. Furthermore, rewards grounded in control theory, such as those based on CLFs, enhance reference tracking (*102, 203*). Other methods embed controllers directly within RL training, for instance using LIP-based planners to provide desired foot placements (*101*). Trajectory computation can also be performed online, where MPC is computed during both training and deployment to enable adaptive motion generation (*204*). Finally, integrating control-theoretic safety notions with RL provides formal safety guarantees (*205, 206*).

Summary of Performance Boundaries of Learning-based Control

Despite advances improving the generalizability, robustness, and sample efficiency of learning-based humanoid locomotion control, several limitations persist. First, these methods remain algorithmically fragile: without motion references or curricula, they often converge to unnatural gaits (*95, 135*) and fail on challenging terrain (*147*). Second, transferring a single policy across tasks or morphologies remains difficult, as embodiment differences and motion retargeting typically require retraining or hierarchical control (*163, 207*). Third, sim-to-real fidelity is still limited. Domain randomization mitigates physics mismatches but cannot fully eliminate them, especially for loco-manipulation and interaction tasks. Existing correction methods are often task-specific or complicate the pipeline (*58, 61, 208*). Fourth, many policies rely solely on proprioception, omitting perception and tactile sensing. Perception improves observability but introduces latency, brittleness, and new sim-to-real gaps (*209*), while added sensing increases calibration and compute demands. Fifth, even with GPU-parallel simulation, training still requires high-end hardware, extensive environment engineering, and nontrivial real-world data collection for system identification and edge-case coverage (*128*). Finally, ensuring safety is still a major challenge: learning-based controllers operate largely as black boxes, making failures unpredictable. Though constrained RL and control-theoretic filters can enhance safety, they remain rare and add modeling and computational overhead (*210, 211, 212, 213*).

Emerging Frontiers in Humanoid Robot Control

Recent advances in RL have expanded humanoid control beyond its classical foundations, paving the way for generative approaches that promise greater adaptability and naturalness in motion (**Fig. 1**). These advances (**Fig. 4**) include a progression from *discriminative* to *generative* modeling, moving from direct, reactive input-output mappings toward models that learn underlying data distributions and generate or predict future trajectories; from *uni-modal* to *multi-modal* learning that unifies perception, language, and proprioception; from *single-task* to *multi-task* control with shared policies for diverse behaviors; and from *locomotion* to *loco-manipulation*, extending control from agile mobility to whole-body, tool-using interaction.

The driving force behind these shifts is the demand for general-purpose humanoids capable of human-level reasoning and physical performance in diverse, complex, novel environments. Breakthroughs in foundation-scale AI tools, such as Large Language Models (LLMs) (*103*), Vision-Language Models (VLMs) (*105*), Vision-Language-Action Models (VLAs) (*183*), diffusion models (*97*), and world models (*108*), make this goal increasingly attainable. The models offer expressive capacity for diverse motion styles, robustness to sparse or partial data,

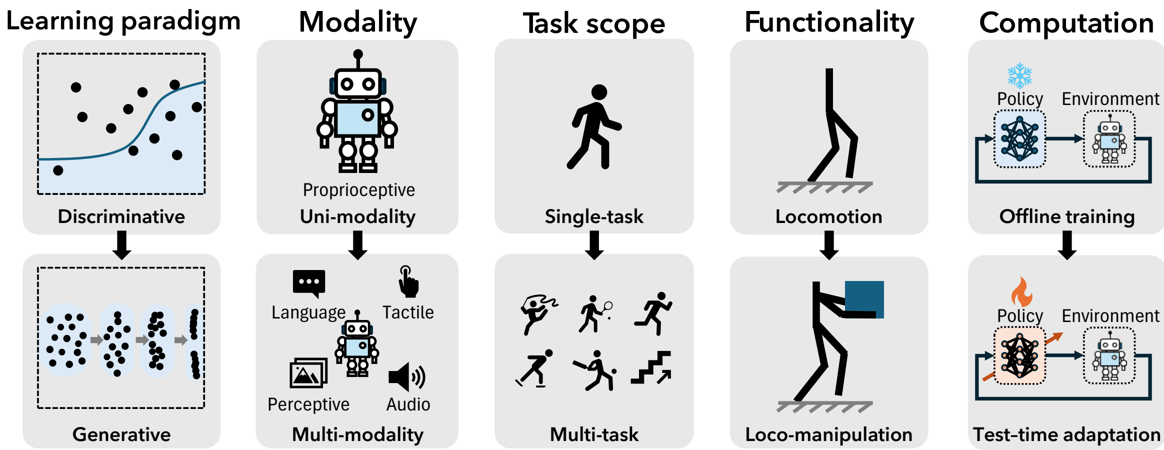

Figure 4: Emerging research shifts in humanoid control across five dimensions. Current trends reflect a transition from discriminative to generative models, uni-modal to multi-modal sensing, single-task to multi-task learning, locomotion to loco-manipulation, and offline training to test-time adaptation. These shifts point toward unified generative approaches that integrate perception, reasoning, and control to enable humanoids to generalize across modalities, tasks, functionalities, and environments.

seamless integration of multi-modality, and enhanced transferability for zero-shot or few-shot adaptation across terrains and morphologies. Examples include next-token prediction for sequential motion generation (*186*), diffusion policies for terrain-adaptive walking (*97*), and multi-modal foundation models for high-level locomotion planning (*105, 214*).

Early progress in other domains, primarily tabletop manipulation (*215*), suggests that generative models could enable humanoids to perform agile whole-body manipulation, mobile tool use, and physically interactive collaboration. Yet, integrating these models into humanoid control is challenging. Key barriers include data scarcity, safety, and the need for high-frequency inference at control rates above hundreds of Hz. Additional barriers specific to humanoids stem from their underactuated, high-dimensional dynamics, stringent stability requirements, and extreme agility demands.

The remainder of this section surveys the emerging generative paradigms, their connections to classical methods, and the open challenges and opportunities shaping the next decade of humanoid control.

Learning Paradigm Shift: From Discriminative to Generative Models

Motivation and limitations. In learning-based locomotion, *discriminative* models learn a direct mapping from states (or observations) to actions, often producing a single deterministic action. While effective on hardware, such policies are limited to behaviors that are explicitly rewarded or demonstrated (*58, 91, 216*), making multi-modal choices (e.g., stepping left or right) and adaptation to data shifts difficult without retraining. These limitations motivate *generative* policies that model a distribution over feasible actions or trajectories, enabling uncertainty awareness and explicit conditioning at inference time (*95, 135, 189, 217, 218, 219, 220, 221, 222, 223*).

Unlike point estimates, a generative control policy π_ϕ with parameters ϕ learns a distribution over feasible trajectories rather than a single deterministic control sequence. It is trained to reproduce expert demonstrations or real-world trajectories while remaining physically plausible, similar to diffusion policies conditioned on state or observation history. This problem can be expressed as:

$$
\begin{aligned}
\underset{\phi}{\text{maximize}} \quad & J(\pi_\phi) = \mathbb{E}_{\tau \sim p_\tau(\cdot|\pi_\phi)} \big[\log p_{\text{data}}(\tau) \; - \; \lambda \, \mathcal{D}\big(\pi_\phi(\tau) \,\big\|\, p_{\text{phys}}(\tau)\big)\big] && \text{(Objective)} \quad \text{(4a)} \\
\text{subject to} \quad & s_{t+1} \sim p(\cdot \mid s_t, a_t), && \text{(Dynamics)} \quad \text{(4b)} \\
& a_t \sim \pi_\phi(\cdot \mid o_{\le t}), && \text{(Generative model)} \quad \text{(4c)} \\
& o_t \sim \mathcal{O}(\cdot \mid s_t), && \text{(Observation model)} \quad \text{(4d)} \\
& s_t \in \mathcal{S}, \; a_t \in \mathcal{A}, \; o_t \in \mathcal{O}, \; (s_t, a_t) \in \mathcal{C}_{\text{phys}}. && \text{(Constraints)} \quad \text{(4e)}
\end{aligned}
$$

Here, $J(\pi_\phi)$ is a data-driven objective that encourages the policy π_ϕ to match the data distribution $p_{\text{data}}(\tau)$ of experts or real-world trajectories while remaining consistent with physical feasibility via the KL regularization term $\mathcal{D}(\pi_\phi \| \, p_{\text{phys}})$ weighted by $\lambda \ge 0$. The term $p_{\text{phys}}(\tau)$ represents a physics-aware prior or simulator-consistent process that enforces dynamic plausibility. The policy π_ϕ models the conditional action distribution given partial observations o_t, typically parameterized as a denoising process refining latent noise into valid actions. The probabilistic dynamics $s_{t+1} \sim p(\cdot \mid s_t, a_t)$ generalize real or learned transition models, and the observation model $o_t \sim \mathcal{O}(\cdot \mid s_t)$ captures partial observability. The constraint $(s_t, a_t) \in \mathcal{C}_{\text{phys}}$ enforces physical feasibility such as contact, friction, torque, and kinematic limits. This formulation reframes motion generation as sampling from a distribution of feasible futures, unifying classical predictive control and generative modeling within a common decision-making framework.

Pathways from discriminative to generative policies. Bridging these paradigms, adversarial imitation learning reveals motion priors and style rewards (*135, 217*), while transformer- and diffusion-based planners enable trajectory-level reasoning and conditional synthesis (*189, 219, 221, 222*). In visuomotor tasks, diffusion policies perform online replanning within learned manifolds (*223*). World models extend this idea by performing control through latent imagination (*54, 108, 194*). In humanoid locomotion, generative motion models (e.g., adversarial or diffusion-based) sample diverse, physically consistent footstep and CoM trajectories under uncertainty, while discriminative controllers handle fast stabilization (*106, 224*). A hierarchical design, where a generative planner proposes motion distributions and a low-level controller executes them, combines reactivity with expressivity (*186, 221, 223, 225, 226*). This integration transforms locomotion from deterministic mapping to generative reasoning, sampling from feasible futures for adaptive, multi-modal control.

Modality Shift: From Uni-modal to Multi-modal Learning Empowered by Foundation AI Models

Motivation and limitations. Most humanoid locomotion controllers remain *uni-modal*, relying solely on proprioceptive feedback, such as joint encoders, inertial measurement units, and contact sensors, without exteroceptive perception. Rooted in early model-based control (*26, 64, 227, 228*), such "blind" controllers achieve real-time balance but lack look-ahead capability. Although recent DRL policies show strong agility and robustness (*58, 128, 129, 198, 229, 230*), their proprioceptive design limits performance in terrain-aware tasks such as stairs or gaps. This motivates a shift toward *multi-modal* learning that integrates vision, depth, or tactile sensing to enhance anticipation and adaptability. Empirical studies show that vision-based locomotion policies can double success rates over blind baselines (*231,232*). Fusion architectures such as LocoTransformer (*233*), VB-Com (*234*), and attention-based encoders (*93, 235*) further improve terrain adaptation and temporal consistency. Overall, uni-modal controllers remain efficient but short-sighted, motivating perception-rich multi-modal frameworks.

From uni-modal control to multi-modal integration. Achieving robust multi-modal control requires architectures that preserve modality-specific cues while supporting cross-modal reasoning. Recent frameworks fuse proprioception, vision, and additional inputs (e.g., lidar and tactile) using structured fusion or Transformer-based cross-attention (*233, 236*), enhancing terrain affordance modeling and contextual adaptability (*215, 237*). Large robot datasets (*184, 193*) and foundation AI models have accelerated this trend. Vision-language-action (VLA) systems such as RT-2 (*184*), π_0 (*238*), and Gemini Robotics (*239*) unify perception, reasoning, and control through internet-scale pre-training. Humanoid-oriented variants like Helix (*225*) and GR00T N1 (*226*) embed vision, proprioception, and language-conditioned goals into unified latent spaces. Such policies enable zero-shot generalization to unseen terrains and semantically grounded locomotion. This marks a shift from reactive stabilization to context-aware, intelligent control for generalist humanoid agents.

Task Scope Shift: From Single-Task to Multi-Task Policy with Generalizability and Adaptability

Motivation and limitations. Conventional humanoid locomotion controllers are optimized for *single tasks*, e.g., walking, running, or jumping, each trained with its own rewards and hyperparameters. While effective within domain, they lack scalability and fail to adapt to new goals or conditions without retraining. This narrow specialization hinders the creation of versatile humanoids capable of composing or adapting skills. Recent work therefore pursues *multi-task* policies that share representations across tasks to learn reusable locomotion priors. Training one model over diverse tasks promotes shared structure, data efficiency, and adaptability beyond rigid single-task setups (*58, 128, 129, 198*). However, multi-task learning introduces reward conflicts, gradient interference, and catastrophic forgetting (*240, 241*), calling for structured training and scalable architectures.

Toward generalizable multi-task policies. Recent approaches seek controllers that generalize across tasks. Multi-task RL uses shared parameters or modular architectures to co-train diverse skills (*242, 243*), while hierarchical schemes organize reusable motion primitives and high-level composition (*244, 245*). Diffusion-based motion synthesis enables skill blending and interpolation at inference (*221, 222*). Cross-embodiment generalization leverages heterogeneous datasets (e.g., BridgeData V2 (*246*), RT-X (*193*)) and meta-learning or distillation (*247*) to share latent skill spaces (*248, 249*). The latest trend, large-scale behavior foundation models (*184,215,225,226,236,237*), unifies task, morphology, and sensing modalities within one latent policy, enabling zero-shot transfer and open-world locomotion. This marks the transition from handcrafted single-skill control to broadly adaptive agents.

Functionality Shift: From Locomotion to Loco-Manipulation and Interaction

Motivation and limitations. Traditional humanoid research has treated *locomotion* and *manipulation* separately: locomotion focuses on balance and terrain adaptation (*26, 58, 128*), while manipulation emphasizes dexterous grasping and contact-rich control (*250, 251, 252*). This separation limits coordinated whole-body behaviors. Locomotion-centric systems can traverse but not interact, while manipulation frameworks often assume a static base and ignore body dynamics. To enable general-purpose humanoids capable of assistance, tool use, and human collaboration, locomotion must evolve into *loco-manipulation*, a unified paradigm jointly reasoning over movement, contact, and force.

Toward unified loco-manipulation and interaction. Recent studies pursue this unification through human motion mimicking, hierarchical control, multi-modal perception, and generative planning. For single tasks, co-tracking approaches jointly track robot, object, and contact references from retargeted human-scene interaction data (*100, 163, 173, 174*). Diffusion-based models synthesize physically plausible human-object interaction trajectories (*253*), while task-and-motion planning couples contact reasoning with long-horizon objectives (*254*). Hierarchical systems integrate high-level vision planning with low-level RL (*136*). Dual-agent RL frameworks such as FALCON (*150, 179*) synchronize upper- and lower-body control, and hybrid visuomotor systems (*255*) combine visual and proprioceptive cues for tasks like carrying or door operation. Unified loco-manipulation controllers such as ULC (*256*) achieve end-to-end walking and handling coordination. At scale, behavior foundation models (e.g., Helix and GR00T N1 (*193, 225, 226*)) integrate locomotion, manipulation, and interaction trajectories into shared visuomotor latent spaces, enabling zero-shot transfer across tasks and morphologies. These advances represent the transition from locomotion-focused control to embodied humanoids that move, manipulate, and collaborate fluidly in complex environments.

Inherent Connections to Classical Control

Despite their seemingly radical departure from traditional paradigms, generative and foundation-scale models share core principles with classical control (**Tab. 2**). LLMs function as dynamic feedback systems: their evolving context window acts as memory (state), updated by each new observation and token to form a closed feedback loop in latent space, as formalized in *Prompt a Robot to Walk* (*103*). Diffusion models (*257*) mirror stochastic optimal control, where denoising solves a stochastic differential equation minimizing an implicit energy functional, analogous to optimal control under uncertainty via the Hamilton-Jacobi-Bellman framework. Representation learning in LLMs (*103*), VLMs (*105*), and VLAs (*183, 226*) likewise seeks latent manifolds where tasks become controllable and observable, echoing classical concepts of state estimation and observability. The transformer's self-attention mechanism can be viewed as adaptive gain scheduling (*258*), dynamically weighting tokens by contextual relevance to implement state-dependent feedback. In-context learning further parallels adaptive control (*44, 199*), where controllers refine online from input-output behavior without explicit parameter updates. Viewed through this lens, generative models extend rather than replace control-theoretic foundations, embodying feedback through autoregression, adaptation through context updates, gain scheduling through attention, and optimality through diffusion guidance.

Toward the Next Generation of Humanoid Control

Humanoid control is evolving from walking and manipulation toward safe, reliable, and useful operation around people. Solving this open challenge now depends on integrating control with safety, scalable hardware, perception, and human-compatible decision making to transform lab systems into deployable collaborators.

Safer and more reliable humanoids. Falling remains the main barrier to humanoid deployment. Standardized tests for recovery, fall rate, and impact severity (*259*), beyond average performance, are needed to evaluate robustness, especially in rare failure cases (*210*). Achieving reliability will rely on compliant actuation, energy-dissipating design, and test-time adaptation (**Fig. 4**) that adjusts control online to unforeseen disturbances during deployment.

Scalable and accessible platforms. The cost of humanoid platforms has dropped by over an order of magnitude through advances in actuators, lightweight design, and scalable manufacturing. Continued reductions via mass production could democratize access, but affordability alone is insufficient. To move from research prototypes to practical, economically viable systems, platforms must also be durable, maintainable, and capable of recovering from minor damage without expert repair.

Perceptual and cognitive intelligence. Next-generation humanoids will integrate control with cognitive and perceptual intelligence. Multi-modal sensing and generative planning will enable anticipatory, context-aware behavior, blurring the boundary between control and cognition so humanoids can act, understand, anticipate, and convey intent in human-compatible ways. This integration naturally suggests a dual-system hierarchy (**Fig. 3B**): a fast, reflexive "System 1" for stabilization and real-time interaction, and a slower, deliberative "System 2" for predictive reasoning (*260, 261*). Achieving effective coordination between these layers may be key to unified cognitive-motor control in future humanoids.

Embodied intelligence ecosystems. Humanoid robots could become as transformative as personal computers. While computers enabled disembodied intelligence in software, humanoids will bring embodied intelligence into the physical world. Their evolution may mirror the path of personal computers: from isolated machines to networked, specialized embodiments ("an internet of humanoids") across contexts. This shift could enable shared platforms, new industries, and research in embodied safety, ethical autonomy, and large-scale human-robot collaboration.

Getting Started in Humanoid Locomotion Control

This section offers concrete guidance on hardware choice, software infrastructure, RL workflow, and foundational references so new teams could build controllers that are physically grounded rather than opaque.

Starting from first principles. Even with learning-based methods, understanding dynamics, feedback control, and optimization remains indispensable. These principles explain gait stability, controller failure, and hardware safety, making learned policies more interpretable and reliable.

Hardware: build or buy. Teams typically choose between open-source and commercial humanoids. Open-source systems, such as the Berkeley Humanoid (*262*), Berkeley Humanoid Lite (*263*), Stanford ToddlerBot (*264*), NimbRo-OP2X (*265*), and Poppy Humanoid (*266*), offer public designs, actuators, and baseline controllers. They are relatively low-cost and modifiable but require careful assembly and maintenance. Commercial robots, including Unitree's G1 and H1, Booster's T1 and K1, EngineAI's SE/PM, PNDbotics' Adam, and U.S. platforms such as Agility's Digit and Apptronik's Apollo, provide safety-tested hardware, integrated software, and vendor support. They cost more but accelerate algorithm development and improve reproducibility. In short, open-source systems maximize flexibility, while commercial systems maximize productivity.

Open-source software: control stacks, simulators, and data. Progress is increasingly enabled by shared infrastructure. For model-based control, MIT Cheetah (*267*) and OpenLoong (*268*) demonstrate whole-body and MPC-based locomotion. CasADi (*269*), Acados (*270*), OCS2 (*271*), and Judo (*272*) make nonlinear predictive control practical. For learning-based control, Isaac Lab (*273*), MuJoCo Playground (*141*), and MJLab (*274*) support GPU-accelerated RL simulation, while Newton (*275*) and Genesis (*276*) scale to large parallel computation with differentiable physics. Human motion datasets such as AMASS (*153*) and pipelines like BeyondMimic (*97, 163*) help encode human-like behaviors and enhance robust transfer to hardware.

Getting started with RL. A practical progression includes: (*a*) Install Isaac Sim and Isaac Lab; (*b*) Execute a walking policy in simulation; (*c*) Train balancing and walking controllers; (*d*) Condition on commanded velocity; (*e*) Track human motion; (*f*) Learn perceptive locomotion on rough terrain; and (*g*) deploy a learned policy from (*c*)-(*f*) on hardware.

Core references. For deeper study, we recommend: *Feedback Control of Dynamic Bipedal Robot Locomotion* (*30*) and *Robot Modeling and Control* (*277*) for dynamics and classical feedback control; *Reinforcement Learning: An Introduction* (*113*) and *PPO* (*91*) for decision making; *DreamerV3* (*194*) for world-model based control; *Deep Learning* (*278*), *Introduction to Variational Autoencoders* (*279*), and *The Principles of Diffusion Models* (*280*) for generative foundations.

Conclusion

This paper has reviewed the evolution of humanoid locomotion control from classical real-time feedback and predictive control to reinforcement learning and emerging generative paradigms. We established a unified perspective that connects these paradigms through shared principles of physics-based modeling, constrained decision-making, and adaptation to uncertainties. We clarified the distinct characteristics of each paradigm: classical control provides stability and interpretability, learning-based approaches enable agility and robustness, and emerging generative methods promise versatility and generalization across tasks and functionalities. Beyond this synthesis, we provided practical recommendations that lower the barrier to entry for new researchers and practitioners in humanoid control. Finally, we outlined future challenges and opportunities toward humanoids capable of intelligent, adaptive, and safe operation in real-world human environments.

References

1. A. Roychoudhury, S. Khorshidi, S. Agrawal, M. Bennewitz, Perception for humanoid robots. *Current Robotics Reports* **4** (4), 127–140 (2023).

2. Y. Tong, H. Liu, Z. Zhang, Advancements in humanoid robots: A comprehensive review and future prospects. *IEEE/CAA Journal of Automatica Sinica* **11** (2), 301–328 (2024).

3. Z. Zhao, S. Cheng, Y. Ding, Z. Zhou, S. Zhang, D. Xu, Y. Zhao, A Survey of Optimization-Based Task and Motion Planning: From Classical to Learning Approaches. *IEEE/ASME Transactions on Mechatronics* **30** (4), 2799–2825 (2025).

4. W. Talbot, J. Nubert, T. Tuna, C. Cadena, F. Dümbgen, J. Tordesillas, T. D. Barfoot, M. Hutter, Continuous-Time State Estimation Methods in Robotics: A Survey. *IEEE Transactions on Robotics* **41**, 4975–4999 (2025).

5. J. Reher, A. D. Ames, Dynamic walking: Toward agile and efficient bipedal robots. *Annual Review of Control, Robotics, and Autonomous Systems* **4** (1), 535–572 (2021).

6. G. Satheesh Kumar, S. Aravind, R. Subramanian, S. K. Bharadwaj, R. R. Muthuraman, R. Steve Mitchell, A. Bucha, M. Sriram, K. Shri Hari, N. J. Robin, Literature survey on four-legged robots. *Trends in Mechanical and Biomedical Design* pp. 691–702 (2021).

7. J. Carpentier, P.-B. Wieber, Recent progress in legged robots locomotion control. *Current Robotics Reports* **2** (3), 231–238 (2021).

8. A. Torres-Pardo, D. Pinto-Fernández, M. Garabini, F. Angelini, D. Rodriguez-Cianca, S. Massardi, J. Tornero, J. C. Moreno, D. Torricelli, Legged locomotion over irregular terrains: State of the art of human and robot performance. *Bioinspiration & Biomimetics* **17** (6), 061002 (2022).

9. T. Mikołajczyk, E. Mikołajewska, H. F. N. Al-Shuka, T. Malinowski, A. Kłodowski, D. Y. Pimenov, T. Paczkowski, F. Hu, K. Giasin, D. Mikołajewski, M. Macko, Recent advances in bipedal walking robots: Review of gait, drive, sensors and control systems. *Sensors* **22** (12), 4440 (2022).

10. G. Picardi, A. Astolfi, D. Chatzievangelou, J. Aguzzi, M. Calisti, Underwater legged robotics: Review and perspectives. *Bioinspiration & Biomimetics* **18** (3), 031001 (2023).

11. Y. Gong, G. Sun, A. Nair, A. Bidwai, R. CS, J. Grezmak, G. Sartoretti, K. A. Daltorio, Legged robots for object manipulation: A review. *Frontiers in Mechanical Engineering* **9**, 1142421 (2023).

12. Y. Zhao, J. Wang, G. Cao, Y. Yuan, X. Yao, L. Qi, Intelligent control of multilegged robot smooth motion: a review. *IEEE Access* **11**, 86645–86685 (2023).

13. S. Ha, J. Lee, M. van de Panne, Z. Xie, W. Yu, M. Khadiv, Learning-based legged locomotion: State of the art and future perspectives. *The International Journal of Robotics Research* **44** (8), 1396–1427 (2025).

14. P. M. Wensing, M. Posa, Y. Hu, A. Escande, N. Mansard, A. Del Prete, Optimization-based control for dynamic legged robots. *IEEE Transactions on Robotics* **40**, 43–63 (2023).

15. Z. Gu, J. Li, W. Shen, W. Yu, Z. Xie, S. McCrory, X. Cheng, A. Shamsah, R. Griffin, C. K. Liu, A. Kheddar, X. B. Peng, Y. Zhu, G. Shi, Q. Nguyen, G. Cheng, H. Gao, Y. Zhao, Humanoid Locomotion and Manipulation: Current Progress and Challenges in Control, Planning, and Learning. *arXiv:2501.02116* (2025).

16. M. H. Raibert, *Legged robots that balance* (MIT press) (1986).

17. Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, K. Fujimura, The intelligent ASIMO: System overview and integration, in *IEEE/RSJ international conference on intelligent robots and systems*, vol. 3 (2002), pp. 2478–2483.

18. K. Sreenath, H.-W. Park, I. Poulakakis, A compliant hybrid zero dynamics controller for stable, efficient and fast bipedal walking on MABEL. *The International Journal of Robotics Research* **30** (9), 1170–1193 (2011).

19. S. Rezazadeh, C. Hubicki, M. Jones, A. Peekema, J. Van Why, A. Abate, J. Hurst, Spring-mass walking with ATRIAS in 3D: Robust gait control spanning zero to 4.3 kph on a heavily underactuated bipedal robot, in *Dynamic Systems and Control Conference*, vol. 57243 (2015), p. V001T04A003.

20. A. Hereid, C. M. Hubicki, E. A. Cousineau, A. D. Ames, Dynamic humanoid locomotion: A scalable formulation for HZD gait optimization. *IEEE Transactions on Robotics* **34** (2), 370–387 (2018).

21. W.-L. Ma, S. Kolathaya, E. R. Ambrose, C. M. Hubicki, A. D. Ames, Bipedal robotic running with DURUS-2D: Bridging the gap between theory and experiment, in *International conference on hybrid systems: computation and control* (2017), pp. 265–274.

22. S. Feng, E. Whitman, X. Xinjilefu, C. G. Atkeson, Optimization-based full body control for the DARPA robotics challenge. *Journal of field robotics* **32** (2), 293–312 (2015).

23. E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, C. Orlowski, The DARPA Robotics Challenge Finals: Results and Perspectives, in *The DARPA robotics challenge finals: Humanoid robots to the rescue*, pp. 1–26 (2018).

24. M. Johnson, B. Shrewsbury, S. Bertrand, T. Wu, D. Duran, M. Floyd, P. Abeles, D. Stephen, N. Mertins, A. Lesman, J. Carff, W. Rifenburgh, P. Kaveti, W. Straatman, J. Smith, M. Griffioen, B. Layton, T. De Boer,

T. Koolen, J. Pratt, Team IHMC's lessons learned from the DARPA robotics challenge trials. *Journal of Field Robotics* **32** (2), 192–208 (2015).

25. R. Today, "Recent progress on ATLAS, the world's most dynamic humanoid robot" - Scott Kuindersma (2020), `https://www.youtube.com/watch?v=EGABAx52GKI`.

26. S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, R. Tedrake, Optimization-based locomotion planning, estimation, and control design for the ATLAS humanoid robot. *Autonomous robots* **40** (3), 429–455 (2016).

27. J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, S. Kim, Dynamic locomotion in the MIT Cheetah 3 through convex model-predictive control, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2018), pp. 1–9.

28. A. J. Ijspeert, Biorobotics: Using robots to emulate and investigate agile locomotion. *Science* **346** (6206), 196–203 (2014).

29. C. Li, F. Qian, Swift progress for robots over complex terrain. *Nature* **616**, 252–253 (2023).

30. E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, B. Morris, *Feedback Control of Dynamic Bipedal Robot Locomotion* (CRC Press) (2007).

31. J. P. Reher, A. Hereid, S. Kolathaya, C. M. Hubicki, A. D. Ames, Algorithmic foundations of realizing multi-contact locomotion on the humanoid robot DURUS, in *Algorithmic Foundations of Robotics XII: Proceedings of the Twelfth Workshop on the Algorithmic Foundations of Robotics* (Springer) (2020), pp. 400–415.

32. R. J. Full, D. E. Koditschek, Templates and anchors: neuromechanical hypotheses of legged locomotion on land. *Journal of experimental biology* **202** (23), 3325–3332 (1999).

33. S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, H. Hirukawa, The 3D linear inverted pendulum mode: A simple modeling for a biped walking pattern generation, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2001), pp. 239–246.

34. C. Hubicki, J. Grimes, M. Jones, D. Renjewski, A. Spröwitz, A. Abate, J. Hurst, ATRIAS: Design and validation of a tether-free 3D-capable spring-mass bipedal robot. *The International Journal of Robotics Research* **35** (12), 1497–1521 (2016).

35. T. McGeer, Passive dynamic walking. *The International Journal of Robotics Research* **9** (2), 62–82 (1990).

36. D. E. Orin, A. Goswami, S.-H. Lee, Centroidal dynamics of a humanoid robot. *Autonomous robots* **35** (2), 161–176 (2013).

37. H. Audren, A. Kheddar, 3-D robust stability polyhedron in multicontact. *IEEE Transactions on Robotics* **34** (2), 388–403 (2018).

38. R. Batke, F. Yu, J. Dao, J. Hurst, R. L. Hatton, A. Fern, K. Green, Optimizing bipedal maneuvers of single rigid-body models for reinforcement learning, in *IEEE-RAS International Conference on Humanoid Robots* (2022), pp. 714–721.

39. X. Xiong, A. Ames, 3-D Underactuated Bipedal Walking via H-LIP Based Gait Synthesis and Stepping Stabilization. *IEEE Transactions on Robotics* **38** (4), 2405–2425 (2022).

40. Y. Gong, J. W. Grizzle, Zero dynamics, pendulum models, and angular momentum in feedback control of bipedal locomotion. *Journal of Dynamic Systems, Measurement, and Control* **144** (12), 121006 (2022).

41. Y. Gao, V. Paredes, Y. Gong, Z. He, A. Hereid, Y. Gu, Time-Varying Foot Placement Control for Humanoid Walking on Swaying Rigid Surface. *IEEE Transactions on Robotics* (2025).

42. G. Mesesan, R. Schuller, J. Englsberger, C. Ott, A. Albu-Schäffer, Unified motion planner for walking, running, and jumping using the three-dimensional divergent component of motion. *IEEE Transactions on Robotics* **39** (6), 4443–4463 (2023).

43. O. Dosunmu-Ogunbi, A. Shrivastava, G. Gibson, J. W. Grizzle, Stair climbing using the angular momentum linear inverted pendulum model and model predictive control, in *Proc. IEEE Int. Conf. Intel. Rob. Syst.* (2023), pp. 8558–8565.

44. J. Stewart, I.-C. Chang, Y. Gu, P. A. Ioannou, Adaptive ankle torque control for bipedal humanoid walking on surfaces with unknown horizontal and vertical motion, in *American Control Conference* (2025), pp. 4647–4652.

45. X. Xiong, A. D. Ames, D. I. Goldman, A stability region criterion for flat-footed bipedal walking on deformable granular terrain, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2017), pp. 4552–4559.

46. A. J. Ijspeert, Central pattern generators for locomotion control in animals and robots: a review. *Neural Networks* **21** (4), 642–653 (2008).

47. B. Han, H. Yi, Z. Xu, X. Yang, X. Luo, 3D-SLIP model based dynamic stability strategy for legged robots with impact disturbance rejection. *Scientific Reports* **12** (1), 5892 (2022).

48. E. Todorov, T. Erez, Y. Tassa, MuJoCo: A physics engine for model-based control, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2012), pp. 5026–5033.

49. V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, G. State, Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv:2108.10470* (2021).

50. J. Hwangbo, J. Lee, M. Hutter, Per-contact iteration method for solving contact dynamics. *IEEE Robotics and Automation Letters* **3** (2), 895–902 (2018).

51. E. Coumans, Y. Bai, Pybullet, a python module for physics simulation for games, robotics and machine learning (2016).

52. P. Roth, J. Frey, C. Cadena, M. Hutter, Learned Perceptive Forward Dynamics Model for Safe and Platform-aware Robotic Navigation. *arXiv:2504.19322* (2025).

53. A. Nagabandi, K. Konolige, S. Levine, V. Kumar, Deep dynamics models for learning dexterous manipulation, in *Conference on Robot Learning* (2020), pp. 1101–1112.

54. D. Hafner, T. Lillicrap, J. Ba, M. Norouzi, Dream to control: Learning behaviors by latent imagination. *arXiv:1912.01603* (2019).

55. Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, J. Morimoto, Deep learning, reinforcement learning, and world models. *Neural Networks* **152**, 267–275 (2022).

56. P. Wu, A. Escontrela, D. Hafner, P. Abbeel, K. Goldberg, Daydreamer: World models for physical robot learning, in *Conference on Robot Learning* (2023), pp. 2226–2240.

57. M. Q. Ali, A. Sridhar, S. Matiana, A. Wong, M. Al-Sharman, Humanoid World Models: Open World Foundation Models for Humanoid Robotics. *arXiv:2506.01182* (2025).

58. J. Hwangbo, J. Lee, A. Dosovitskiy, C. D. Bellicoso, V. Tsounis, V. Koltun, M. Hutter, Learning Agile and Dynamic Motor Skills for Legged Robots. *Science Robotics* **4** (26), eaau5872 (2019).

59. A. H. Chang, C. M. Hubicki, J. J. Aguilar, D. I. Goldman, A. D. Ames, P. A. Vela, Learning to jump in granular media: Unifying optimal control synthesis with Gaussian process-based regression, in *IEEE International Conference on Robotics and Automation* (2017), pp. 2154–2160.

60. W.-S. Yang, W.-C. Lu, P.-C. Lin, Legged robot running using a physics-data hybrid motion template. *IEEE Transactions on Robotics* **37** (5), 1680–1695 (2021).

61. T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan, Z. Yi, G. Qu, K. Kitani, J. Hodgins, L. J. Fan, Y. Zhu, C. Liu, G. Shi, ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills. *arXiv:2502.01143* (2025).

62. M. Vukobratović, D. Stokić, Dynamic control of unstable locomotion robots. *Mathematical Biosciences* **24** (1-2), 129–157 (1975).

63. S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, H. Hirukawa, Biped walking pattern generation by using preview control of the zero-moment point, in *IEEE International Conference on Robotics and Automation*, vol. 2 (2003), pp. 1620–1626.

64. J. Pratt, J. Carff, S. Drakunov, A. Goswami, Capture Point: A Step toward Humanoid Push Recovery, in *IEEE-RAS International Conference on Humanoid Robots* (2006), pp. 200–207.

65. X. Xiong, A. D. Ames, Dynamic and Versatile Humanoid Walking via Embedding 3D Actuated SLIP Model With Hybrid LIP Based Stepping. *IEEE Robotics and Automation Letters* **5** (4), 6286–6293 (2020).

66. M. Mistry, J. Buchli, S. Schaal, Inverse dynamics control of floating base systems using orthogonal decomposition, in *IEEE international conference on robotics and automation* (2010), pp. 3406–3412.

67. J. Reher, A. D. Ames, Inverse dynamics control of compliant hybrid zero dynamic walking, in *IEEE International Conference on Robotics and Automation* (2021), pp. 2040–2047.

68. H. Sadeghian, C. Ott, G. Garofalo, G. Cheng, Passivity-based control of underactuated biped robots within hybrid zero dynamics approach, in *IEEE International Conference on Robotics and Automation* (2017), pp. 4096–4101.

69. I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, K. Sreenath, Real-world humanoid locomotion with reinforcement learning. *Science Robotics* **9** (89), eadi9579 (2024).

70. E. Westervelt, J. Grizzle, D. Koditschek, Hybrid zero dynamics of planar biped walkers. *IEEE Transactions on Automatic Control* **48** (1), 42–56 (2003).

71. A. D. Ames, K. Galloway, K. Sreenath, J. W. Grizzle, Rapidly exponentially stabilizing control lyapunov functions and hybrid zero dynamics. *IEEE Transactions on Automatic Control* **59** (4), 876–891 (2014).

72. K. Galloway, K. Sreenath, A. D. Ames, J. W. Grizzle, Torque saturation in bipedal robotic walking through control Lyapunov function-based quadratic programs. *IEEE Access* **3**, 323–332 (2015).

73. C. Dario Bellicoso, F. Jenelten, P. Fankhauser, C. Gehring, J. Hwangbo, M. Hutter, Dynamic locomotion and whole-body control for quadrupedal robots, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2017), pp. 3359–3365.

74. A. W. Winkler, C. D. Bellicoso, M. Hutter, J. Buchli, Gait and trajectory optimization for legged systems through phase-based end-effector parameterization. *IEEE Robotics and Automation Letters* **3** (3), 1560–1567 (2018).

75. K. Bouyarmane, S. Caron, A. Escande, A. Kheddar, Multi-contact motion planning and control, in *Humanoid Robotics: A Reference* (Springer), pp. 1763–1804 (2019).

76. A. D. Ames, Human-inspired control of bipedal walking robots. *IEEE Transactions on Automatic Control* **59** (5), 1115–1130 (2014).

77. C. Khazoom, S. Hong, M. Chignoli, E. Stanger-Jones, S. Kim, Tailoring solution accuracy for fast whole-body model predictive control of legged robots. *IEEE Robotics and Automation Letters* (2024).

78. A. B. Ghansah, S. A. Esteban, A. D. Ames, Hierarchical Reduced-Order Model Predictive Control for Robust Locomotion on Humanoid Robots, in *International Conference on Humanoid Robots* (2025), pp. 1–8.

79. G. Romualdi, S. Dafarra, G. L'Erario, I. Sorrentino, S. Traversaro, D. Pucci, Online Non-linear Centroidal MPC for Humanoid Robot Locomotion with Step Adjustment, in *International Conference on Robotics and Automation* (2022), pp. 10412–10419.

80. R. Grandia, F. Jenelten, S. Yang, F. Farshidian, M. Hutter, Perceptive locomotion through nonlinear model-predictive control. *IEEE Transactions on Robotics* **39** (5), 3402–3421 (2023).

81. G. Kim, D. Kang, J.-H. Kim, S. Hong, H.-W. Park, Contact-implicit Model Predictive Control: Controlling diverse quadruped motions without pre-planned contact modes or trajectories. *The International Journal of Robotics Research* **44** (3), 486–510 (2025).

82. S. Le Cleac'h, T. A. Howell, S. Yang, C.-Y. Lee, J. Zhang, A. Bishop, M. Schwager, Z. Manchester, Fast Contact-Implicit Model-Predictive Control. *IEEE Transactions on Robotics* **40**, 1617–1629 (2024).

83. V. Kurtz, A. Castro, A. Ö. Önol, H. Lin, Inverse Dynamics Trajectory Optimization for Contact-Implicit Model Predictive Control. *International Journal of Robotics Research* **44** (3), 320–348 (2025).

84. N. J. Kong, C. Li, G. Council, A. M. Johnson, Hybrid iLQR Model Predictive Control for Contact Implicit Stabilization on Legged Robots. *IEEE Transactions on Robotics* **39** (6), 4712–4727 (2023).

85. H. Xue, C. Pan, Z. Yi, G. Qu, G. Shi, Full-order sampling-based mpc for torque-level locomotion control via diffusion-style annealing. *arXiv:2409.15610* (2024).

86. T. Howell, N. Gileadi, S. Tunyasuvunakool, K. Zakka, T. Erez, Y. Tassa, Predictive sampling: Real-time behaviour synthesis with MuJoCo. *arXiv:2212.00541* (2022).

87. J. Alvarez-Padilla, J. Z. Zhang, S. Kwok, J. M. Dolan, Z. Manchester, Real-time whole-body control of legged robots with model-predictive path integral control, in *IEEE International Conference on Robotics and Automation* (2025), pp. 14721–14727.

88. K. Bouyarmane, K. Chappellet, J. Vaillant, A. Kheddar, Quadratic programming for multirobot and task-space force control. *IEEE Transactions on Robotics* **35** (1), 64–77 (2018).

89. R. Deits, R. Tedrake, Footstep planning on uneven terrain with mixed-integer convex optimization, in *IEEE-RAS international conference on humanoid robots* (2014), pp. 279–286.

90. J. Schulman, S. Levine, P. Abbeel, M. Jordan, P. Moritz, Trust region policy optimization, in *International Conference on Machine Learning* (2015).

91. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. *arXiv:1707.06347* (2017).

92. Z. Xie, H. Y. Ling, N. H. Kim, M. Van De Panne, Allsteps: curriculum-driven learning of stepping stone skills, in *Computer Graphics Forum*, vol. 39 (2020), pp. 213–224.

93. T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, M. Hutter, Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science robotics* **7** (62), eabk2822 (2022).

94. T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, J. Humplik, M. Wulfmeier, S. Tunyasuvunakool, N. Y. Siegel, R. Hafner, M. Bloesch, K. Hartikainen, A. Byravan, L. Hasenclever, Y. Tassa, F. Sadeghi, N. Batchelor, F. Casarini, S. Saliceti, C. Game, N. Sreendra, K. Patel, M. Gwira, A. Huber, N. Hurley, F. Nori, R. Hadsell, N. Heess, Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *Science Robotics* **9** (89), eadi8022 (2024).

95. X. B. Peng, P. Abbeel, S. Levine, M. van de Panne, DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills. *ACM Transactions on Graphics* **37** (4), 143:1–143:14 (2018).

96. T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. M. Kitani, C. Liu, G. Shi, OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning, in *Conference on Robot Learning* (2024).

97. T. E. Truong, Q. Liao, X. Huang, G. Tevet, C. K. Liu, K. Sreenath, BeyondMimic: From Motion Tracking to Versatile Humanoid Control via Guided Diffusion. *arXiv:2508.08241* (2025).

98. K. Green, Y. Godse, J. Dao, R. L. Hatton, A. Fern, J. Hurst, Learning Spring Mass Locomotion: Guiding Policies With a Reduced-Order Model. *IEEE Robotics and Automation Letters* **6** (2), 3926–3932 (2021).

99. Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, K. Sreenath, Reinforcement Learning for Robust Parameterized Locomotion Control of Bipedal Robots, in *IEEE International Conference on Robotics and Automation* (2021), pp. 2811–2817.

100. F. Liu, Z. Gu, Y. Cai, Z. Zhou, H. Jung, J. Jang, S. Zhao, S. Ha, Y. Chen, D. Xu, Y. Zhao, Opt2skill: Imitating dynamically-feasible whole-body trajectories for versatile humanoid loco-manipulation. *IEEE Robotics and Automation Letters* (2025).

101. H. J. Lee, S. Hong, S. Kim, Integrating Model-Based Footstep Planning with Model-Free Reinforcement Learning for Dynamic Legged Locomotion, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2024), pp. 11248–11255.

102. K. Li, Z. Olkin, Y. Yue, A. D. Ames, CLF-RL: Control Lyapunov Function Guided Reinforcement Learning. *arXiv:2508.09354* (2025).

103. Y.-J. Wang, B. Zhang, J. Chen, K. Sreenath, Prompt a robot to walk with large language models, in *IEEE Conference on Decision and Control* (2024), pp. 1531–1538.

104. Y. Shao, X. Huang, B. Zhang, Q. Liao, Y. Gao, Y. Chi, Z. Li, S. Shao, K. Sreenath, LangWBC: Language-directed Humanoid Whole-Body Control via End-to-end Learning. *arXiv:2504.21738* (2025).

105. Y. Guo, Y.-J. Wang, L. Zha, J. Chen, DoReMi: Grounding language model by detecting and recovering from plan-execution misalignment, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2024), pp. 12124–12131.

106. G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, A. H. Bermano, Human Motion Diffusion Model, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023).

107. X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, J. Chen, Advancing Humanoid Locomotion: Mastering Challenging Terrains with Denoising World Model Learning, in *Robotics: Science and Systems (RSS)* (2024).

108. H. Liu, Y. Gao, S. Teng, Y. Chi, Y. S. Shao, Z. Li, M. Ghaffari, K. Sreenath, Ego-Vision World Model for Humanoid Contact Planning. *arXiv:2510.11682* (2025).

109. A. G. Barto, R. S. Sutton, C. W. Anderson, Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-13** (5), 834–846 (1983).

110. C. J. C. H. Watkins, P. Dayan, Q-learning. *Machine Learning* **8** (3–4), 279–292 (1992).

111. R. J. Williams, Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Machine Learning* **8** (3–4), 229–256 (1992).

112. D. P. Bertsekas, J. N. Tsitsiklis, *Neuro-Dynamic Programming* (Athena Scientific) (1996).

113. R. S. Sutton, A. G. Barto, *Reinforcement learning: An introduction*, vol. 1 (MIT press Cambridge) (1998).

114. J. N. Tsitsiklis, B. Van Roy, An Analysis of Temporal-Difference Learning with Function Approximation. *IEEE Transactions on Automatic Control* **42** (5), 674–690 (1997).

115. L. C. Baird, Residual Algorithms: Reinforcement Learning with Function Approximation, in *International Conference on Machine Learning* (1995), pp. 30–37.

116. J. Kober, J. A. Bagnell, J. Peters, Reinforcement Learning in Robotics: A Survey. *The International Journal of Robotics Research* **32** (11), 1238–1274 (2013).

117. M. P. Deisenroth, G. Neumann, J. Peters, A Survey on Policy Search for Robotics. *Foundations and Trends in Robotics* **2** (1–2), 1–142 (2013).

118. G. Tesauro, Temporal Difference Learning and TD-Gammon. *Communications of the ACM* **38** (3), 58–68 (1995).

119. L.-J. Lin, Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning* **8** (3–4), 293–321 (1992).

120. M. Riedmiller, Neural fitted Q iteration—first experiences with a data efficient neural reinforcement learning method, in *European Conference on Machine Learning* (2005), pp. 317–328.

121. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems*, vol. 25 (2012), pp. 1097–1105.

122. Y. LeCun, Y. Bengio, G. Hinton, Deep Learning. *Nature* **521**, 436–444 (2015).

123. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning. *Nature* **518** (7540), 529–533 (2015).

124. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing Atari with Deep Reinforcement Learning. *arXiv:1312.5602* (2013).

125. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search. *Nature* **529** (7587), 484–489 (2016).

126. A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, N. Dormann, Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research* **22** (268), 1–8 (2021).

127. E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, I. Stoica, RLlib: Abstractions for Distributed Reinforcement Learning, in *International Conference on Machine Learning* (2018), pp. 3053–3062.

128. N. Rudin, D. Hoeller, P. Reist, M. Hutter, Learning to walk in minutes using massively parallel deep reinforcement learning, in *Conference on Robot Learning* (2022), pp. 91–100.

129. J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, V. Vanhoucke, Sim-to-Real: Learning Agile Locomotion For Quadruped Robots. *Robotics: Science and Systems XIV* (2018).

130. X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, S. Levine, Learning agile robotic locomotion skills by imitating animals. *arXiv:2004.00784* (2020).

131. W. Yu, V. C. Kumar, G. Turk, C. K. Liu, Sim-to-real transfer for biped locomotion, in *international conference on intelligent robots and systems* (2019), pp. 3503–3510.

132. Z. Xie, P. Clary, J. Dao, P. Morais, J. Hurst, M. van de Panne, Learning Locomotion Skills for Cassie: Iterative Design and Sim-to-Real, in *Conference on Robot Learning* (2020), pp. 317–329.

133. H. Duan, A. Malik, J. Dao, A. Saxena, K. Green, J. Siekmann, A. Fern, J. Hurst, Sim-to-real learning of footstep-constrained bipedal dynamic walking, in *International Conference on Robotics and Automation* (2022), pp. 10428–10434.

134. F. Shi, Y. Kojio, T. Makabe, T. Anzai, K. Kojima, K. Okada, M. Inaba, Reference-free learning bipedal motor skills via assistive force curricula, in *The International Symposium of Robotics Research* (Springer) (2022), pp. 304–320.

135. X. B. Peng, Z. Ma, P. Abbeel, S. Levine, A. Kanazawa, Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (ToG)* **40** (4), 1–20 (2021).

136. T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, G. Shi, Learning human-to-humanoid real-time whole-body teleoperation, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2024), pp. 8944–8951.

137. M. Ji, X. Peng, F. Liu, J. Li, G. Yang, X. Cheng, X. Wang, Exbody2: Advanced expressive humanoid whole-body control. *arXiv:2412.13196* (2024).

138. Z. Su, B. Zhang, N. Rahmanian, Y. Gao, Q. Liao, C. Regan, K. Sreenath, S. S. Sastry, HITTER: A humanoid table tennis robot via hierarchical planning and learning. *arXiv:2508.21043* (2025).

139. M. Hu, W. Chen, W. Li, F. Mandali, Z. He, R. Zhang, P. Krisna, K. Christian, L. Benaharon, D. Ma, K. Ramani, Y. Gu, Towards Versatile Humanoid Table Tennis: Unified Reinforcement Learning with Prediction Augmentation. *arXiv:2509.21690* (2025).

140. L. P. Kaelbling, M. L. Littman, A. R. Cassandra, Planning and acting in partially observable stochastic domains. *Artificial intelligence* **101** (1-2), 99–134 (1998).

141. K. Zakka, B. Tabanpour, Q. Liao, M. Haiderbhai, S. Holt, J. Y. Luo, A. Allshire, E. Frey, K. Sreenath, L. A. Kahrs, C. Sferrazza, Y. Tassa, P. Abbeel, MuJoCo Playground. *arXiv:2502.08844* (2025).

142. A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, Stable Baselines, `https://github.com/hill-a/stable-baselines` (2018).

143. G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, OpenAI Gym. *arXiv:1606.01540* (2016).

144. J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, Domain randomization for transferring deep neural networks from simulation to the real world, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2017), pp. 23–30.

145. Z. Xie, X. Da, M. Van de Panne, B. Babich, A. Garg, Dynamics randomization revisited: A case study for quadrupedal locomotion, in *IEEE International Conference on Robotics and Automation* (2021), pp. 4955–4961.

146. Y. Bengio, J. Louradour, R. Collobert, J. Weston, Curriculum learning, in *International Conference on Machine Learning* (2009), pp. 41–48.

147. J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, M. Hutter, Learning quadrupedal locomotion over challenging terrain. *Science robotics* **5** (47), eabc5986 (2020).

148. D. Chen, B. Zhou, V. Koltun, P. Krähenbühl, Learning by cheating, in *Conference on Robot Learning* (2020), pp. 66–75.

149. L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, P. Abbeel, Asymmetric Actor Critic for Image-Based Robot Learning, in *Proceedings of Robotics: Science and Systems* (2018).

150. Y. Li, Y. Zhang, W. Xiao, C. Pan, H. Weng, G. He, T. He, G. Shi, Hold My Beer: Learning Gentle Humanoid Locomotion and End-Effector Stabilization Control, in *Conference on Robot Learning* (2025).

151. Y. Xue, W. Dong, M. Liuˆ, W. Zhang, J. Pang, A Unified and General Humanoid Whole-Body Controller for Versatile Locomotion. *arXiv:2502.03206* (2025).

152. H. Zhang, L. Zhang, Z. Chen, L. Chen, Y. Wang, R. Xiong, Natural Humanoid Robot Locomotion with Generative Motion Prior. *arXiv:2503.09015* (2025).

153. N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, M. J. Black, AMASS: Archive of motion capture as surface shapes, in *IEEE/CVF international conference on computer vision* (2019), pp. 5442–5451.

154. C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence* **36** (7), 1325–1339 (2013).

155. Carnegie Mellon University Graphics Lab, CMU Graphics Lab Motion Capture Database, `http://mocap.cs.cmu.edu/`.

156. C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, T. Asfour, Unifying Representations and Large–Scale Whole-Body Motion Databases for Studying Human Motion. *IEEE Transactions on Robotics* **32** (4), 796–809 (2016).

157. S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, N. F. Troje, MoVi: A Large Multipurpose Motion and Video Dataset. *arXiv:2003.01888* (2020).

158. C. Guo, S. Zou, X. Zuo, S. Wang, W. Ji, X. Li, L. Cheng, Generating Diverse and Natural 3D Human Motions From Text, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 5152–5161.

159. J. Li, J. Wu, C. K. Liu, Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)* **42** (6), 1–11 (2023).

160. J. Lin, A. Zeng, S. Lu, Y. Cai, R. Zhang, H. Wang, L. Zhang, Motion-X: A large-scale 3D expressive whole-body human motion dataset. *Advances in Neural Information Processing Systems* **36**, 25268–25280 (2023).

161. Z. Luo, J. Cao, K. Kitani, W. Xu, Perpetual humanoid control for real-time simulated avatars, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 10895–10904.

162. L. Penco, B. Clément, V. Modugno, E. M. Hoffman, G. Nava, D. Pucci, N. G. Tsagarakis, J.-B. Mouret, S. Ivaldi, Robust real-time whole-body motion retargeting from human to humanoid, in *IEEE-RAS 18th International Conference on Humanoid Robots* (2018), pp. 425–432.

163. L. Yang, X. Huang, Z. Wu, A. Kanazawa, P. Abbeel, C. Sferrazza, C. K. Liu, R. Duan, G. Shi, OmniRetarget: Interaction-Preserving Data Generation for Humanoid Whole-Body Loco-Manipulation and Scene Interaction. *arXiv:2509.26633* (2025).

164. J. P. Araujo, Y. Ze, P. Xu, J. Wu, C. K. Liu, Retargeting Matters: General Motion Retargeting for Humanoid Motion Tracking. *arXiv:2510.02252* (2025).

165. J. Merel, Y. Tassa, D. TB, S. Srinivasan, J. Lemmon, Z. Wang, G. Wayne, N. Heess, Learning human behaviors from motion capture by adversarial imitation. *arXiv:1707.02201* (2017).

166. A. Tang, T. Hiraoka, N. Hiraoka, F. Shi, K. Kawaharazuka, K. Kojima, K. Okada, M. Inaba, HumanMimic: Learning natural locomotion and transitions for humanoid robot via wasserstein adversarial imitation, in *IEEE International Conference on Robotics and Automation* (2024), pp. 13107–13114.

167. Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, C. Finn, HumanPlus: Humanoid Shadowing and Imitation from Humans, in *Conference on Robot Learning* (2025), pp. 2828–2844.

168. X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, X. Wang, Expressive whole-body control for humanoid robots. *arXiv:2402.16796* (2024).

169. Z. Chen, M. Ji, X. Cheng, X. Peng, X. B. Peng, X. Wang, GMT: General Motion Tracking for Humanoid Whole-Body Control. *arXiv:2506.14770* (2025).

170. Y. Ze, Z. Chen, J. P. AraÃšjo, Z.-a. Cao, X. B. Peng, J. Wu, C. K. Liu, TWIST: Teleoperated Whole-Body Imitation System. *arXiv:2505.02833* (2025).

171. K. Yin, W. Zeng, K. Fan, M. Dai, Z. Wang, Q. Zhang, Z. Tian, J. Wang, J. Pang, W. Zhang, UniTracker: Learning universal whole-body motion tracker for humanoid robots. *arXiv:2507.07356* (2025).

172. T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang, L. Fan, Y. Zhu, Hover: Versatile neural whole-body controller for humanoid robots, in *IEEE International Conference on Robotics and Automation* (2025), pp. 9989–9996.

173. H. Weng, Y. Li, N. Sobanbabu, Z. Wang, Z. Luo, T. He, D. Ramanan, G. Shi, HDMI: Learning Interactive Humanoid Whole-Body Control from Human Videos. *arXiv:2509.16757* (2025).

174. S. Zhao, Y. Ze, Y. Wang, C. K. Liu, P. Abbeel, G. Shi, R. Duan, ResMimic: From General Motion Tracking to Humanoid Whole-body Loco-Manipulation via Residual Learning. *arXiv:2510.05070* (2025).

175. N. Sobanbabu, G. He, T. He, Y. Yang, G. Shi, Sampling-based System Identification with Active Exploration for Legged Sim2Real Learning, in *9th Annual Conference on Robot Learning* (2025).

176. D. Hoeller, N. Rudin, D. Sako, M. Hutter, ANYmal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics* **9** (88), eadi7566 (2024).

177. Y. Wang, M. Yang, W. Zeng, Y. Zhang, X. Xu, H. Jiang, Z. Ding, Z. Lu, From Experts to a Generalist: Toward General Whole-Body Control for Humanoid Robots. *arXiv:2506.12779* (2025).

178. N. Fey, G. B. Margolis, M. Peticco, P. Agrawal, Bridging the Sim-to-Real Gap for Athletic Loco-Manipulation. *arXiv:2502.10894* (2025).

179. Y. Zhang, Y. Yuan, P. Gurunath, T. He, S. Omidshafiei, A. akbar Agha-mohammadi, M. Vazquez-Chanlatte, L. Pedersen, G. Shi, FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation. *arXiv:2505.06776* (2025).

180. D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, J. Davidson, Learning latent dynamics for planning from pixels, in *International Conference on Machine Learning* (2019), pp. 2555–2565.

181. G. A. Castillo, B. Weng, W. Zhang, A. Hereid, Data-driven latent space representation for robust bipedal locomotion learning, in *IEEE International Conference on Robotics and Automation* (2024), pp. 1172–1178.

182. X. Huang, T. Truong, Y. Zhang, F. Yu, J. P. Sleiman, J. Hodgins, K. Sreenath, F. Farshidian, Diffuse-CLoC: Guided diffusion for physics-based character look-ahead control. *ACM Transactions on Graphics* **44** (4), 1–12 (2025).

183. H. Xue, X. Huang, D. Niu, Q. Liao, T. Kragerud, J. T. Gravdahl, X. B. Peng, G. Shi, T. Darrell, K. Sreenath, S. Sastry, LeVERB: Humanoid Whole-Body Control with Latent Vision-Language Instruction. *arXiv:2506.13751* (2025).

184. B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, Q. Vuong, V. Vanhoucke, H. Tran, R. Soricut, A. Singh, J. Singh, P. Sermanet, P. R. Sanketi, G. Salazar, M. S. Ryoo, K. Reymann, K. Rao, K. Pertsch, I. Mordatch, H. Michalewski, Y. Lu, S. Levine, L. Lee, T.-W. E. Lee, I. Leal, Y. Kuang, D. Kalashnikov, R. Julian, N. J. Joshi, A. Irpan, B. Ichter, J. Hsu, A. Herzog, K. Hausman, K. Gopalakrishnan, C. Fu, P. Florence, C. Finn, K. A. Dubey, D. Driess, T. Ding, K. M. Choromanski, X. Chen, Y. Chebotar, J. Carbajal, N. Brown, A. Brohan, M. G. Arenas, K. Han, RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control, in *Proceedings of The 7th Conference on Robot Learning*, vol. 229 of *Proceedings of Machine Learning Research* (2023), pp. 2165–2183.

185. H. Liu, S. Teng, B. Liu, W. Zhang, M. Ghaffari, Discrete-Time Hybrid Automata Learning: Legged Locomotion Meets Skateboarding. *arXiv:2503.01842* (2025).

186. I. Radosavovic, B. Zhang, B. Shi, J. Rajasegaran, S. Kamat, T. Darrell, K. Sreenath, J. Malik, Humanoid locomotion as next token prediction. *Advances in neural information processing systems* **37**, 79307–79324 (2024).

187. G. Feng, H. Zhang, Z. Li, X. B. Peng, B. Basireddy, L. Yue, Z. Song, L. Yang, Y. Liu, K. Sreenath, S. Levine, GenLoco: Generalized locomotion controllers for quadrupedal robots, in *Conference on Robot Learning* (2023), pp. 1893–1903.

188. E. Chane-Sane, P.-A. Leziart, T. Flayols, O. Stasse, P. Souères, N. Mansard, CaT: Constraints as Terminations for Legged Locomotion Reinforcement Learning, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2024), pp. 13303–13310.

189. Z. Wang, J. J. Hunt, M. Zhou, Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning. *arXiv:2208.06193* (2022).

190. M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan, K. Hausman, A. Herzog, D. Ho, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, E. Jang, R. J. Ruano, K. Jeffrey, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, K.-H. Lee, S. Levine, Y. Lu, L. Luu, C. Parada, P. Pastor,

J. Quiambao, K. Rao, J. Rettinghouse, D. Reyes, P. Sermanet, N. Sievers, C. Tan, A. Toshev, V. Vanhoucke, F. Xia, T. Xiao, P. Xu, S. Xu, M. Yan, A. Zeng, Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. *arXiv:2204.01691* (2022).

191. K. Chen, X. Huang, X. Chen, J. Yi, Contingency model predictive control for bipedal locomotion on moving surfaces with a linear inverted pendulum model, in *2024 American Control Conference (ACC)* (IEEE) (2024), pp. 3166–3171.

192. Q. Nguyen, K. Sreenath, Robust safety-critical control for dynamic robotics. *IEEE Transactions on Automatic Control* **67** (3), 1073–1088 (2021).

193. E. Collaboration, A. O'Neill, A. Rehman, A. Gupta, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, A. Tung, A. Bewley, A. Herzog, A. Irpan, A. Khazatsky, A. Rai, A. Gupta, A. Wang, A. Kolobov, A. Singh, A. Garg, A. Kembhavi, A. Xie, A. Brohan, A. Raffin, A. Sharma, A. Yavary, A. Jain, A. Balakrishna, A. Wahid, B. Burgess-Limerick, B. Kim, B. Schölkopf, B. Wulfe, B. Ichter, C. Lu, C. Xu, C. Le, C. Finn, C. Wang, C. Xu, C. Chi, C. Huang, C. Chan, C. Agia, C. Pan, C. Fu, C. Devin, D. Xu, D. Morton, D. Driess, D. Chen, D. Pathak, D. Shah, D. Büchler, D. Jayaraman, D. Kalashnikov, D. Sadigh, E. Johns, E. Foster, F. Liu, F. Ceola, F. Xia, F. Zhao, F. V. Frujeri, F. Stulp, G. Zhou, G. S. Sukhatme, G. Salhotra, G. Yan, G. Feng, G. Schiavi, G. Berseth, G. Kahn, G. Yang, G. Wang, H. Su, H.-S. Fang, H. Shi, H. Bao, H. B. Amor, H. I. Christensen, H. Furuta, H. Bharadhwaj, H. Walke, H. Fang, H. Ha, I. Mordatch, I. Radosavovic, I. Leal, J. Liang, J. Abou-Chakra, J. Kim, J. Drake, J. Peters, J. Schneider, J. Hsu, J. Vakil, J. Bohg, J. Bingham, J. Wu, J. Gao, J. Hu, J. Wu, J. Wu, J. Sun, J. Luo, J. Gu, J. Tan, J. Oh, J. Wu, J. Lu, J. Yang, J. Malik, J. Silvério, J. Hejna, J. Booher, J. Tompson, J. Yang, J. Salvador, J. J. Lim, J. Han, K. Wang, K. Rao, K. Pertsch, K. Hausman, K. Go, K. Gopalakrishnan, K. Goldberg, K. Byrne, K. Oslund, K. Kawaharazuka, K. Black, K. Lin, K. Zhang, K. Ehsani, K. Lekkala, K. Ellis, K. Rana, K. Srinivasan, K. Fang, K. P. Singh, K.-H. Zeng, K. Hatch, K. Hsu, L. Itti, L. Y. Chen, L. Pinto, L. Fei-Fei, L. Tan, L. J. Fan, L. Ott, L. Lee, L. Weihs, M. Chen, M. Lepert, M. Memmel, M. Tomizuka, M. Itkina, M. G. Castro, M. Spero, M. Du, M. Ahn, M. C. Yip, M. Zhang, M. Ding, M. Heo, M. K. Srirama, M. Sharma, M. J. Kim, M. Z. Irshad, N. Kanazawa, N. Hansen, N. Heess, N. J. Joshi, N. Suenderhauf, N. Liu, N. D. Palo, N. M. M. Shafiullah, O. Mees, O. Kroemer, O. Bastani, P. R. Sanketi, P. T. Miller, P. Yin, P. Wohlhart, P. Xu, P. D. Fagan, P. Mitrano, P. Sermanet, P. Abbeel, P. Sundaresan, Q. Chen, Q. Vuong, R. Rafailov, R. Tian, R. Doshi, R. Martín-Martín, R. Baijal, R. Scalise, R. Hendrix, R. Lin, R. Qian, R. Zhang, R. Mendonca,

R. Shah, R. Hoque, R. Julian, S. Bustamante, S. Kirmani, S. Levine, S. Lin, S. Moore, S. Bahl, S. Dass, S. Sonawani, S. Tulsiani, S. Song, S. Xu, S. Haldar, S. Karamcheti, S. Adebola, S. Guist, S. Nasiriany, S. Schaal, S. Welker, S. Tian, S. Ramamoorthy, S. Dasari, S. Belkhale, S. Park, S. Nair, S. Mirchandani, T. Osa, T. Gupta, T. Harada, T. Matsushima, T. Xiao, T. Kollar, T. Yu, T. Ding, T. Davchev, T. Z. Zhao, T. Armstrong, T. Darrell, T. Chung, V. Jain, V. Kumar, V. Vanhoucke, V. Guizilini, W. Zhan, W. Zhou, W. Burgard, X. Chen, X. Chen, X. Wang, X. Zhu, X. Geng, X. Liu, X. Liangwei, X. Li, Y. Pang, Y. Lu, Y. J. Ma, Y. Kim, Y. Chebotar, Y. Zhou, Y. Zhu, Y. Wu, Y. Xu, Y. Wang, Y. Bisk, Y. Dou, Y. Cho, Y. Lee, Y. Cui, Y. Cao, Y.-H. Wu, Y. Tang, Y. Zhu, Y. Zhang, Y. Jiang, Y. Li, Y. Li, Y. Iwasawa, Y. Matsuo, Z. Ma, Z. Xu, Z. J. Cui, Z. Zhang, Z. Fu, Z. Lin, Open X-Embodiment: Robotic Learning Datasets and RT-X Models (2025), `https://arxiv.org/abs/2310.08864`.

194. D. Hafner, J. Pasukonis, J. Ba, T. Lillicrap, Mastering Diverse Domains through World Models. *arXiv:2301.04104* (2023).

195. M. Dai, X. Xiong, J. Lee, A. D. Ames, Data-driven adaptation for robust bipedal locomotion with step-to-step dynamics, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2023), pp. 8574–8581.

196. A. Kumar, Z. Li, J. Zeng, D. Pathak, K. Sreenath, J. Malik, Adapting rapid motor adaptation for bipedal robots, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2022), pp. 1161–1168.

197. K. Zhou, J. C. Doyle, *Essentials of robust control*, vol. 104 (Prentice hall Upper Saddle River, NJ) (1998).

198. A. Kumar, Z. Fu, D. Pathak, J. Malik, RMA: Rapid Motor Adaptation for Legged Robots, in *Robotics: Science and Systems* (2021).

199. P. A. Ioannou, J. Sun, *Robust adaptive control*, vol. 1 (PTR Prentice-Hall Upper Saddle River, NJ) (1996).

200. I. T. Jolliffe, J. Cadima, Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374** (2065), 20150202 (2016).

201. G. H. Golub, C. Reinsch, Singular value decomposition and least squares solutions, in *Linear algebra* (Springer), pp. 134–151 (1971).

202. C. Zhang, W. Xiao, T. He, G. Shi, WoCoCo: Learning Whole-Body Humanoid Control with Sequential Contacts, in *Conference on Robot Learning* (2025), pp. 455–472.

203. Z. Olkin, K. Li, W. D. Compton, A. D. Ames, Chasing Stability: Humanoid Running via Control Lyapunov Function Guided Reinforcement Learning. *arXiv:2509.19573* (2025).

204. F. Jenelten, J. He, F. Farshidian, M. Hutter, DTC: Deep Tracking Control. *Science Robotics* **9** (86), eadh5401 (2024).

205. T. He, C. Zhang, W. Xiao, G. He, C. Liu, G. Shi, Agile but safe: Learning collision-free high-speed legged locomotion. *arXiv:2401.17583* (2024).

206. L. Yang, B. Werner, M. d. S. A. D. Ames, CBF-RL: Safety Filtering Reinforcement Learning in Training with Control Barrier Functions. *arXiv:2510.14959* (2025).

207. C. Sferrazza, D.-M. Huang, X. Lin, Y. Lee, P. Abbeel, HumanoidBench: Simulated humanoid benchmark for whole-body locomotion and manipulation. *arXiv:2403.10506* (2024).

208. P. Huang, X. Zhang, Z. Cao, S. Liu, M. Xu, W. Ding, J. Francis, B. Chen, D. Zhao, What went wrong? closing the sim-to-real gap via differentiable causal discovery, in *Conference on Robot Learning* (2023), pp. 734–760.

209. T. Miki, J. Lee, L. Wellhausen, E. Kaufmann, J. Hwangbo, C. D. Bellicoso, F. Tresoldi, M. Fässler, M. Hutter, Perceptive locomotion in rough terrain. *Science Robotics* **7** (66), eabk2822 (2022).

210. F. Shi, C. Zhang, T. Miki, J. Lee, M. Hutter, S. Coros, Rethinking Robustness Assessment: Adversarial Attacks on Learning-based Quadrupedal Locomotion Controllers, in *Robotics: Science and Systems* (2024).

211. J. Achiam, D. Held, A. Tamar, P. Abbeel, Constrained policy optimization, in *International conference on machine learning* (2017), pp. 22–31.

212. J. K. Wabersich, M. N. Zeilinger, A Predictive Safety Filter for Learning-Based Control of Constrained Nonlinear Dynamical Systems. *Automatica* **129**, 109597 (2021).

213. A. D. Ames, S. Xu, J. W. Grizzle, P. Tabuada, Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control* **62** (8), 3861–3876 (2017).

214. A.-C. Cheng, Y. Ji, Z. Yang, Z. Gongye, X. Zou, J. Kautz, E. Bıyık, H. Yin, S. Liu, X. Wang, NaVILA: Legged robot vision-language-action model for navigation. *arXiv:2412.04453* (2024).

215. R. Firoozi, J. Tucker, S. Tian, A. Majumdar, J. Sun, W. Liu, Y. Zhu, S. Song, A. Kapoor, K. Hausman, B. Ichter, D. Driess, J. Wu, C. Lu, M. Schwager, Foundation Models in Robotics: Applications, Challenges, and the Future. *The International Journal of Robotics Research* **44** (5), 701–739 (2024).

216. T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, in *International Conference on Machine Learning* (2018), pp. 1861–1870.

217. J. Ho, S. Ermon, Generative Adversarial Imitation Learning, in *Advances in Neural Information Processing Systems* (2016).

218. Y. Li, J. Song, S. Ermon, InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations, in *Advances in Neural Information Processing Systems*, vol. 30 (2017).

219. L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, I. Mordatch, Decision Transformer: Reinforcement Learning via Sequence Modeling, in *Advances in Neural Information Processing Systems*, vol. 34 (2021), pp. 15084–15097.

220. M. Janner, Q. C. Li, S. Levine, Offline Reinforcement Learning as One Big Sequence Modeling Problem, in *Advances in Neural Information Processing Systems*, vol. 34 (2021), pp. 1273–1286.

221. M. Janner, G. Varun, A. Ajay, J. Tenenbaum, S. Levine, Planning with Diffusion for Flexible Behavior Synthesis, in *International Conference on Machine Learning* (2022).

222. A. Ajay, Y. Du, A. Gupta, J. Tenenbaum, T. Jaakkola, P. Agrawal, Is Conditional Generative Modeling All You Need for Decision-Making? *arXiv:2211.15657* (2022).

223. C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, S. Song, Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. *The International Journal of Robotics Research* **44** (10-11), 1684–1704 (2024).

224. Y. Yuan, J. Song, U. Iqbal, A. Vahdat, J. Kautz, PhysDiff: Physics-Guided Human Motion Diffusion Model. *arXiv:2212.02500* (2022).

225. F. AI, Helix: A Vision-Language-Action Model for Generalist Humanoid Control (2025), `https://www.figure.ai/news/helix`.

226. NVIDIA, J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. J. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, J. Jang, Z. Jiang, J. Kautz, K. Kundalia, L. Lao, Z. Li, Z. Lin, K. Lin, G. Liu, E. Llontop, L. Magne, A. Mandlekar, A. Narayan, S. Nasiriany, S. Reed, Y. L. Tan, G. Wang, Z. Wang, J. Wang, Q. Wang, J. Xiang, Y. Xie, Y. Xu, Z. Xu, S. Ye, Z. Yu, A. Zhang, H. Zhang, Y. Zhao, R. Zheng, Y. Zhu, GR00T N1: An open foundation model for generalist humanoid robots. *arXiv:2503.14734* (2025).

227. T. Koolen, T. De Boer, J. Rebula, A. Goswami, J. Pratt, Capturability-based analysis and control of legged locomotion, Part 1: Theory and application to three simple gait models. *The international journal of robotics research* **31** (9), 1094–1113 (2012).

228. A. Herzog, N. Rotella, S. Mason, F. Grimminger, S. Schaal, L. Righetti, Momentum Control with Hierarchical Inverse Dynamics on a Torque-controlled Humanoid. *Autonomous Robots* **40** (3), 473–491 (2016).

229. Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, K. Sreenath, Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control. *The International Journal of Robotics Research* **44** (5), 840–888 (2025).

230. G. B. Margolis, G. Yang, K. Paigwar, T. Chen, P. Agrawal, Rapid locomotion via reinforcement learning. *The International Journal of Robotics Research* **43** (4), 572–587 (2024).

231. H. Duan, B. Pandit, M. S. Gadde, B. Van Marum, J. Dao, C. Kim, A. Fern, Learning vision-based bipedal locomotion for challenging terrain, in *IEEE International Conference on Robotics and Automation* (2024), pp. 56–62.

232. A. Loquercio, A. Kumar, J. Malik, Learning Visual Locomotion with Cross-Modal Supervision, in *IEEE International Conference on Robotics and Automation* (2023), pp. 7295–7302.

233. R. Yang, M. Zhang, N. Hansen, H. Xu, X. Wang, Learning Vision-Guided Quadrupedal Locomotion End-to-End with Cross-Modal Transformers, in *International Conference on Learning Representations* (2022).

234. J. Ren, T. Huang, H. Wang, Z. Wang, Q. Ben, J. Long, Y. Yang, J. Pang, P. Luo, VB-Com: Learning vision-blind composite humanoid locomotion against deficient perception. *arXiv:2502.14814* (2025).

235. J. He, C. Zhang, F. Jenelten, R. Grandia, M. Bächer, M. Hutter, Attention-based map encoding for learning generalized legged locomotion. *Science Robotics* **10** (105), eadv3604 (2025).

236. X. Han, M. Li, J. Chen, K. Zhou, H. Wu, Multimodal Fusion and Vision-Language Models: A Survey for Robot Vision. *Information Fusion* **86**, 1–25 (2025).

237. W. Zhao, K. Gangaraju, F. Yuan, Multimodal Perception-Driven Decision-Making for Human-Robot Interaction: A Survey. *Frontiers in Robotics and AI* **12**, 101234 (2025).

238. K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, S. Jakubczak, T. Jones, L. Ke, S. Levine, A. Li-Bell, M. Mothukuri, S. Nair, K. Pertsch, L. X. Shi, J. Tanner, Q. Vuong, A. Walling, H. Wang, U. Zhilinsky, π0: A vision-language-action flow model for general robot control. *arXiv.2410.24164* (2024).

239. G. R. Team, S. Abeyruwan, J. Ainslie, J.-B. Alayrac, M. G. Arenas, T. Armstrong, A. Balakrishna, R. Baruch, M. Bauza, M. Blokzijl, S. Bohez, K. Bousmalis, A. Brohan, T. Buschmann, A. Byravan, S. Cabi, K. Caluwaerts, F. Casarini, O. Chang, J. E. Chen, X. Chen, H.-T. L. Chiang, K. Choromanski, D. D'Ambrosio, S. Dasari, T. Davchev, C. Devin, N. D. Palo, T. Ding, A. Dostmohamed, D. Driess, Y. Du, D. Dwibedi, M. Elabd, C. Fantacci, C. Fong, E. Frey, C. Fu, M. Giustina, K. Gopalakrishnan, L. Graesser, L. Hasenclever, N. Heess, B. Hernaez, A. Herzog, R. A. Hofer, J. Humplik, A. Iscen, M. G. Jacob, D. Jain, R. Julian, D. Kalashnikov, M. E. Karagozler, S. Karp, C. Kew, J. Kirkland, S. Kirmani, Y. Kuang, T. Lampe, A. Laurens, I. Leal, A. X. Lee, T.-W. E. Lee, J. Liang, Y. Lin, S. Maddineni, A. Majumdar, A. H. Michaely, R. Moreno, M. Neunert, F. Nori, C. Parada, E. Parisotto, P. Pastor, A. Pooley, K. Rao, K. Reymann, D. Sadigh, S. Saliceti, P. Sanketi, P. Sermanet, D. Shah, M. Sharma, K. Shea, C. Shu, V. Sindhwani, S. Singh, R. Soricut, J. T. Springenberg, R. Sterneck, R. Surdulescu, J. Tan, J. Tompson, V. Vanhoucke, J. Varley, G. Vesom, G. Vezzani, O. Vinyals, A. Wahid, S. Welker, P. Wohlhart, F. Xia, T. Xiao, A. Xie, J. Xie, P. Xu, S. Xu, Y. Xu, Z. Xu, Y. Yang, R. Yao, S. Yaroshenko, W. Yu, W. Yuan, J. Zhang, T. Zhang, A. Zhou, Y. Zhou, Gemini robotics: Bringing ai into the physical world. *arXiv:2503.20020* (2025).

240. E. Parisotto, F. Song, J. Rae, R. Pascanu, C. Gulcehre, S. Jayakumar, M. Jaderberg, R. L. Kaufman, A. Clark, S. Noury, M. Botvinick, N. Heess, R. Hadsell, Stabilizing transformers for reinforcement learning, in *International conference on machine learning* (2020), pp. 7487–7498.

241. Y. W. Teh, V. Bapst, W. M. Czarnecki, J. Quan, J. Kirkpatrick, R. Hadsell, N. Heess, R. Pascanu, Distral: Robust Multi-Task Reinforcement Learning, in *Advances in Neural Information Processing Systems* (2017).

242. T. Yu, C. Finn, S. Dasari, A. Xie, T. Zhang, P. Abbeel, S. Levine, One-Shot Imitation from Observing Humans via Domain-Adaptive Meta-Learning, in *Robotics: Science and Systems* (2018).

243. D. Kalashnikov, J. Varley, Y. Chebotar, B. Swanson, R. Jonschkowski, C. Finn, S. Levine, K. Hausman, MT-Opt: Continuous multi-task robotic reinforcement learning at scale. *arXiv:2104.08212* (2021).

244. J. Merel, L. Hasenclever, A. Galashov, A. Ahuja, V. Pham, G. Wayne, Y. Teh, N. Heess, Neural probabilistic motor primitives for humanoid control, in *International Conference on Learning Representations* (2019).

245. L. Pan, Z. Yang, Z. Dou, W. Wang, B. Huang, B. Dai, T. Komura, J. Wang, Tokenhsi: Unified synthesis of physical human-scene interactions through task tokenization, in *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 5379–5391.

246. H. R. Walke, K. Black, T. Z. Zhao, Q. Vuong, C. Zheng, P. Hansen-Estruch, A. W. He, V. Myers, M. J. Kim, M. Du, A. Lee, K. Fang, C. Finn, S. Levine, BridgeData V2: A dataset for robot learning at scale, in *Conference on Robot Learning* (2023), pp. 1723–1736.

247. C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in *International Conference on Machine Learning* (2017), pp. 1126–1135.

248. K. Rakelly, A. Zhou, C. Finn, S. Levine, D. Quillen, Efficient off-policy meta-reinforcement learning via probabilistic context variables, in *International conference on machine learning* (2019), pp. 5331–5340.

249. T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, S. Levine, Meta-World: A benchmark and evaluation for multi-task and meta reinforcement learning, in *Conference on robot learning* (2020), pp. 1094–1100.

250. OpenAI, M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, W. Zaremba, Learning dexterous in-hand manipulation. *The International Journal of Robotics Research* **39** (1), 3–20 (2020).

251. A. Rajeswaran, V. Kumar, A. Gupta, E. Todorov, S. Levine, Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations, in *Robotics: Science and Systems* (2018).

252. A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, T. Jackson, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, K.-H. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada,

J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. Ryoo, G. Salazar, P. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, B. Zitkovich, RT-1: Robotics transformer for real-world control at scale. *arXiv:2212.06817* (2022).

253. I. Taouil, H. Zhao, A. Dai, M. Khadiv, Physically Consistent Humanoid Loco-Manipulation using Latent Diffusion Models. *arXiv:2504.16843* (2025).

254. M. Ciebielski, V. Dhédin, M. Khadiv, Task and Motion Planning for Humanoid Loco-Manipulation, in *IEEE-RAS 24th International Conference on Humanoid Robots* (2025), pp. 1179–1186.

255. C. Wang, Q. Huang, X. Chen, Z. Zhang, J. Shi, Robust Visuomotor Control for Humanoid Loco-Manipulation Using Hybrid Reinforcement Learning. *Biomimetics* **10** (7), 469 (2025).

256. W. Sun, L. Feng, B. Cao, Y. Liu, Y. Jin, Z. Xie, ULC: A Unified and Fine-Grained Controller for Humanoid Loco-Manipulation. *arXiv:2507.06905* (2025).

257. J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, in *Advances in Neural Information Processing Systems*, vol. 33 (2020), pp. 6840–6851.

258. W. J. Rugh, J. S. Shamma, Research on gain scheduling. *Automatica* **36** (10), 1401–1425 (2000).

259. B. Weng, G. A. Castillo, Y.-S. Kang, A. Hereid, Towards standardized disturbance rejection testing of legged robot locomotion with linear impactor: A preliminary study, observations, and implications, in *IEEE International Conference on Robotics and Automation* (2024), pp. 9946–9952.

260. D. Kahneman, Thinking, fast and slow. *Farrar, Straus and Giroux* (2011).

261. M. Zhu, Y. Zhu, J. Li, J. Wen, Z. Xu, Z. Che, C. Shen, Y. Peng, D. Liu, F. Feng, J. Tang, Language-conditioned robotic manipulation with fast and slow thinking, in *IEEE International Conference on Robotics and Automation* (2024), pp. 4333–4339.

262. Q. Liao, B. Zhang, X. Huang, X. Huang, Z. Li, K. Sreenath, Berkeley Humanoid: A research platform for learning-based control, in *IEEE International Conference on Robotics and Automation* (2025), pp. 2897–2904.

263. Y. Chi, Q. Liao, J. Long, X. Huang, S. Shao, B. Nikolic, Z. Li, K. Sreenath, Demonstrating Berkeley Humanoid Lite: An Open-source, Accessible, and Customizable 3D-printed Humanoid Robot. *arXiv:2504.17249* (2025).

264. H. Shi, W. Wang, S. Song, C. K. Liu, ToddlerBot: Open-Source ML-Compatible Humanoid Platform for Loco-Manipulation. *arXiv:2502.00893* (2025).

265. G. Ficht, H. Farazi, A. Brandenburger, D. Rodriguez, D. Pavlichenko, P. Allgeuer, M. Hosseini, S. Behnke, NimbRo-OP2X: Adult-sized open-source 3D printed humanoid robot, in *IEEE-RAS 18th International Conference on Humanoid Robots* (2018), pp. 1–9.

266. Poppy Project: Open-source 3D-printed humanoid platform, `https://www.poppy-project.org/`.

267. MIT Cheetah Software, `https://github.com/mit-biomimetics/Cheetah-Software`, retrieved 2025.

268. L. Humanoid Robot (Shanghai) Co., OpenLoong-DynamicsControl: Motion control framework of humanoid robot based on MPC and WBC (2024), `https://github.com/loongOpen/Openloong-dyn-control.git`.

269. J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, M. Diehl, CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* **11** (1), 1–36 (2019).

270. R. Verschueren, G. Frison, D. Kouzoupis, J. Frey, N. v. Duijkeren, A. Zanelli, B. Novoselnik, T. Albin, R. Quirynen, M. Diehl, acados—a modular open-source framework for fast embedded optimal control. *Mathematical Programming Computation* **14** (1), 147–183 (2022).

271. G. Neunert, T. Stäuble, M. Giftthaler, M. Stäuble, M. R. Tognon, M. Frigerio, J. Buchli, OCS2: An open source library for optimal control of switched systems. *IEEE Transactions on Control Systems Technology* **28** (5), 1313–1325 (2020), `https://github.com/leggedrobotics/ocs2`.

272. A. H. Li, B. Hung, A. D. Ames, J. Wang, S. L. Cleac'h, P. Culbertson, Judo: A User-Friendly Open-Source Package for Sampling-Based Model Predictive Control. *arXiv:2506.17184* (2025).

273. NVIDIA Isaac Lab, `https://developer.nvidia.com/isaac/lab`, retrieved 2025.

274. mujocolab, MJLab: Isaac Lab API, powered by MuJoCo-Warp (2025), `https://github.com/mujocolab/mjlab`.

275. D. R. Newton Physics Engine Contributors (NVIDIA, Google DeepMind, Newton: GPU-Accelerated Physics Simulation Engine for Robotics (2025), `https://github.com/newton-physics/newton`.

276. G. Authors, Genesis: A Generative and Universal Physics Engine for Robotics and Beyond (2024), `https://github.com/Genesis-Embodied-AI/Genesis`.

277. M. W. Spong, S. Hutchinson, M. Vidyasagar, *Robot Modeling and Control* (John Wiley & Sons) (2006).

278. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT press) (2016).

279. D. P. Kingma, M. Welling, *An Introduction to Variational Autoencoders* (Foundations and Trends in Machine Learning) (2019).

280. C.-H. Lai, Y. Song, D. Kim, Y. Mitsufuji, S. Ermon, The Principles of Diffusion Models. *arXiv:2510.21890* (2025).

Acknowledgments

Artificial intelligence tools (ChatGPT) were used during the preparation of this manuscript to refine the writing after the authors completed the initial draft. All content was subsequently reviewed and verified by all authors to ensure technical accuracy and faithful representation of the intended messages. In addition, ChatGPT was employed to assist in generating and modifying sub-figures in **Fig. 1** to maintain a consistent visual style throughout the figure.

Funding: Y.G. and I.C.C. were supported in part by NSF CMMI-2421768, 2423239, and ONR N00014-24-1-2028. H.S. was supported in part by NSF GCR 2524088, 2344956, 1944655 and NIH R01NS141171 and 1R01EB035404. F.S. was supported in part by MOE 24-1234-P0001 and NUS Presidential Young Professorship. Y.J.W. and K.S. were supported by NSF CMMI-1944722 and the Robotics and AI Institute.

Author contributions: Y.G., G.S., F.S., A.D.A., J. Z., H.S., and K.S. conceptualized this paper; Y.G., G.S., F.S., I.C.C., Y.J.W., Q.C., Z.O., I.L.S., Y.F., and H.S. wrote the paper; and Y.G., G.S., F.S., I.C.C., Y.J.W., Q.C., Z.O., I.L.S., Y.F., J.Z., A.D.A., H.S., and K.S. reviewed and edited the paper.

Competing interests: G.S. holds concurrent appointments as an Assistant Professor at Carnegie Mellon University and as an Amazon Scholar. This paper describes work performed at Carnegie Mellon University and is not associated with Amazon. This work was done when J. Zhang was with Apple Inc. and is not associated with Meta Platforms Inc.