Article

A Comprehensive Survey of Deep Reinforcement Learning Techniques for Soft Mobile Robots

Mohammad Hadi Ghodsi [1], Hamed Shahbazi [2,*] and Keivan Torabi [1]

[1] Department of Mechanical Engineering, University of Isfahan, Isfahan 81746-73441, Iran; mh.ghodsi@eng.ui.ac.ir (M.H.G.); torabi@eng.ui.ac.ir (K.T.)

[2] Department of Mechatronic Engineering, University of Isfahan, Isfahan 81746-73441, Iran

* Corresponding author. E-mail: shahbazi@eng.ui.ac.ir (H.S.)

ABSTRACT: Soft robotics has emerged as a promising direction for enabling safe, adaptive, and energy-efficient interactions with unstructured environments due to its inherent compliance. Recently, Deep Reinforcement Learning (DRL) has become a powerful tool for autonomous behavior generation in soft robots, surpassing limitations of classical model-based control. However, despite rapid growth of publications in this domain, there is still a lack of systematic comparative surveys that clarify how different DRL approaches have been used for soft mobile robots, what types of tasks they address, and what performance evaluation criteria have been used. In this article, we review and classify existing works in DRL-enabled soft robotics, focusing particularly on soft mobile systems, and present a structured synthesis of contributions, algorithms, training strategies, and real-world applications. Unlike previous reviews that discuss soft robotics or DRL separately, this paper explicitly provides cross-comparison across DRL paradigms and soft robot tasks, enabling researchers to identify suitable DRL approaches for different soft mobile robotic behaviors. Finally, major challenges and promising future directions are proposed to advance this interdisciplinary research area.

Keywords: Deep reinforcement learning; Soft robotics

1. Introduction

Soft robotics has grown as an important area of research and development because it provides capabilities that rigid body robots generally lack, such as intrinsic compliance, safe physical interaction, deformability, and adaptation to uncertain environments. In parallel, Deep Reinforcement Learning (DRL) has demonstrated significant success in discovering autonomous control policies directly from experience without requiring explicit dynamic models. As soft robots contain extremely high-dimensional, nonlinear, and history-dependent dynamics, DRL becomes especially attractive for enabling locomotion, manipulation, and environmental interaction without the need for handcrafted controllers.

Recent years have shown an increasing number of works combining DRL algorithms and soft robots in locomotion, manipulation, perception, and adaptive control tasks. Nevertheless, although both DRL and soft robots have been reviewed independently in the literature, there is no comprehensive comparative survey focusing specifically on DRL for soft mobile robots, summarizing and comparing learning techniques, representative tasks, and evaluation metrics. Therefore, this work fills this gap by providing a structured comparative analysis of the learning strategies used in existing DRL-enabled soft mobile robotic systems and comparing different learning paradigms across tasks such as locomotion, navigation, interaction, and underwater motion.

The contributions of this survey are as follows:

- We summarize key DRL approaches that have been used in soft mobile robotics.
- We classify and compare existing literature based on robot morphology, DRL algorithm class, and task.
- We provide a comparative synthesis by connecting soft robotic behaviors (such as open-terrain locomotion, underwater swimming, and continuum manipulation) with the DRL paradigms that enabled them, thus giving a direct mapping between tasks and learning techniques.
- We identify current challenges and provide potential future research directions.

2. The Review of Literature on Soft Mobile Robots and DRL

Although several surveys have reviewed soft robotics [1,2] and others have summarized reinforcement learning in robotics [3,4], there is still no focused comparative survey specifically dedicated to the integration of Deep Reinforcement Learning (DRL) in soft mobile robots. Existing works either emphasize biology-inspired soft material technologies, or they analyze DRL algorithms primarily in rigid robotic platforms. This gap motivates the present survey, which consolidates current efforts, compares methodologies, and identifies the key research trends for DRL-driven soft mobile locomotion.

Ijspeert and colleagues in 2007, in [5], presented a spinal cord model and its implementation in an amphibious salamander robot. They demonstrated how a primitive neural circuit for swimming can be extended using new physiogenetic limb oscillatory centers to explain the salamander's ability to switch between swimming and walking, and vice versa. This model proposes neural mechanisms for adjusting speed, direction, and gait type, which are suitable for all quadrupeds. The model also predicts that the limb oscillatory centers have a lower intrinsic frequency than the body oscillatory centers, a prediction that is supported by the presented biological data. CPG-based (Central Pattern Generator) control helps reduce the dimensionality of the motion control problem while maintaining high flexibility to continuously adjust speed, direction, and gait type according to the environmental context (Figure 1).

In 2008, he reviewed the research conducted on locomotor central pattern generators (CPGs), meaning neural circuits capable of generating coordinated patterns of rhythmic output signals with high dimensionality while receiving only simple, low-dimensional input signals. His review initially includes neurobiological observations related to locomotor CPGs and their numerical modeling, focusing on vertebrates. Then, the use of CPG models for controlling the locomotion of articulated robots and specific motion states is examined. He has analyzed the advantages and disadvantages of using CPGs with robots [6].
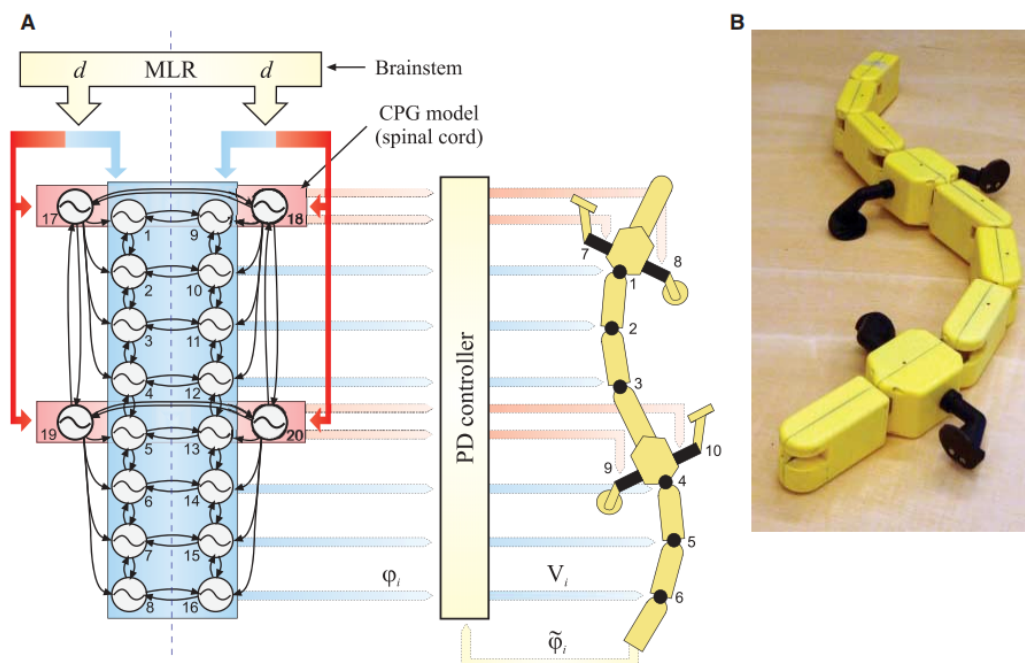

Figure 1. Configuration of the CPG model and salamander robot(**A**) and salamander robot (**B**) [5].

Ijspeert et al. in 2013, in [7], presented and reviewed the assumptions of Dynamic Movement Primitives (DMPs), which are a series of studies focused on modeling the attractor behaviors of autonomous nonlinear dynamical systems using statistical learning techniques. This general design principle is especially useful for modeling motor behaviors in robotics, as well as for modeling biological phenomena. The nature of their approach is such that it starts with a simple dynamical system, such as a set of linear differential equations, and transforms them into a weakly nonlinear system with prescribed attractor dynamics by means of a learnable forcing term. The nonlinear forcing term can be expressed as an independent coupling term that can be learned using standard machine learning techniques that are linear in the parameters. They explain the design principle of their approach and evaluate its properties in several practical examples

in motor control and robotics. In Figure 1, the configuration of the CPG model and the salamander robot is shown.

Jens Kober and his colleagues, in [3], attempted to strengthen the ties between the two research communities by reviewing the work on deep learning for behavior generation in robots. They highlighted both the key challenges in robot reinforcement learning and the notable successes. Furthermore, they studied the role of algorithms, representations, and prior knowledge in achieving these successes. As a result, the special focus of this research is on the selection between model-based and model-free methods, as well as between value function-based and policy search methods. By analyzing a simple problem in detail, they showed how reinforcement learning approaches can be profitably applied.

Daniela Rus and Michael T. Tolley, in [1], discussed recent developments in the field of soft robotics. They also reviewed the design and construction of robots, actuators, stretchable electronic components, sensing systems, power sources, computation and control, modeling and kinematics, and applications of soft robots. In Figure 2, examples of mobile systems inspired by the motion of soft-bodied creatures are shown.
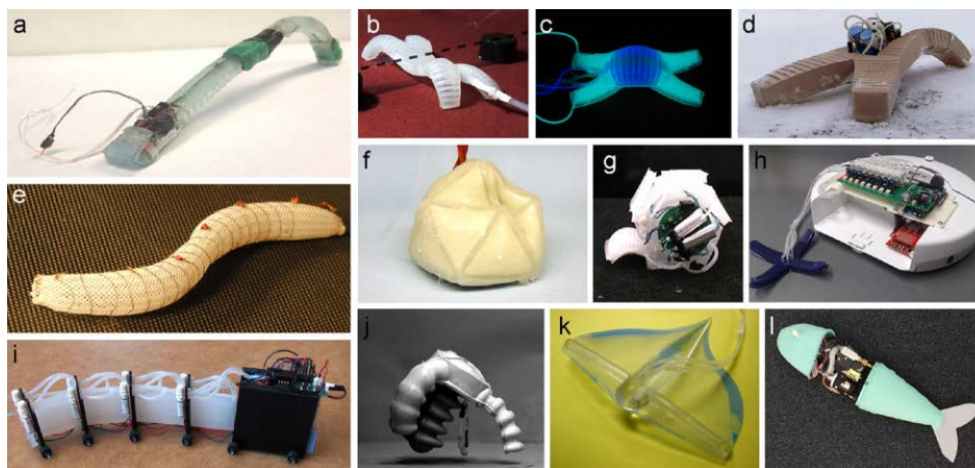
Figure 2. Mobile soft robotic systems inspired by a range of biological systems, demonstrating (**a**) caterpillar-inspired locomotion, (**b**) a multi-gait quadruped, (**c**) active camouflage, (**d**) walking in hazardous environments, (**e**) worm-inspired locomotion, (**f**) particle jamming-based actuation, (**g**) rolling powered by a pneumatic battery, (**h**) a hybrid hard/soft robot, (**i**) snake inspired locomotion, (**j**) jumping powered by internal combustion, (**k**) manta-ray inspired locomotion, and (**l**) an autonomous fish [1].

Wang and colleagues, in [8], proposed learning from demonstration (LfD) for motion planning of arms based on ionic polymer-metal composite (IPMC) actuators using statistical machine learning algorithms. To encode demonstrated trajectories and estimate suitable paths for the arms to reproduce tasks, models are built based on Gaussian Mixture Models (GMM) and Gaussian Mixture Regression (GMR), respectively. In this study, forward and inverse kinematic models of the soft robotic arm based on IPMC were obtained for motion control. A flexible and soft robotic arm (Figure 3) with six IPMC segments was employed and validated the learned trajectories by successfully navigating through a narrow keyhole.
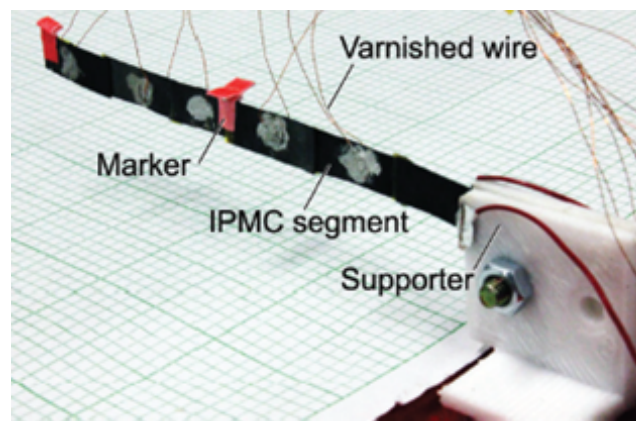
Figure 3. Structure of the flexible IPMC arm [8].

This work proposes the introduction of the learning from demonstration (LfD) method for motion planning of a flexible and soft robotic arm based on IPMC for tasks such as passing through a hole. A skillful flexible IPMC arm with six segments was implemented to validate the learned paths from demonstration. In experiments, this arm was able to autonomously pass through a narrow opening (5 mm). With the current results, the IPMC arm could be used in complex tasks such as grasping tissue while navigating through a body cavity.

Bern and colleagues, in [9], presented a class of tendon-actuated soft robots and ensured their low-cost and usability for non-expert users. The fabrication techniques they introduced are mainly based on traditional methods used for making plush toys (Figure 4). Movement of a plush robot is achieved by driving internal winches that pull (or push) tendons routed beneath the skin. They provided a forward simulation model to predict the deformation behavior of a plush robot given certain contractions of internal winches and used this in an interactive control scheme. They then explained how to invert this forward model to be used in an interactive control design. They also built a hexapod and a gripper to demonstrate the applicability of their system. In this design, the user selects a target position for the robot, and the optimal contractions of the robot's winches are automatically computed in real-time.

(a) laser cut pattern   (b) sew, invert and stuff   (c) rig with tendons
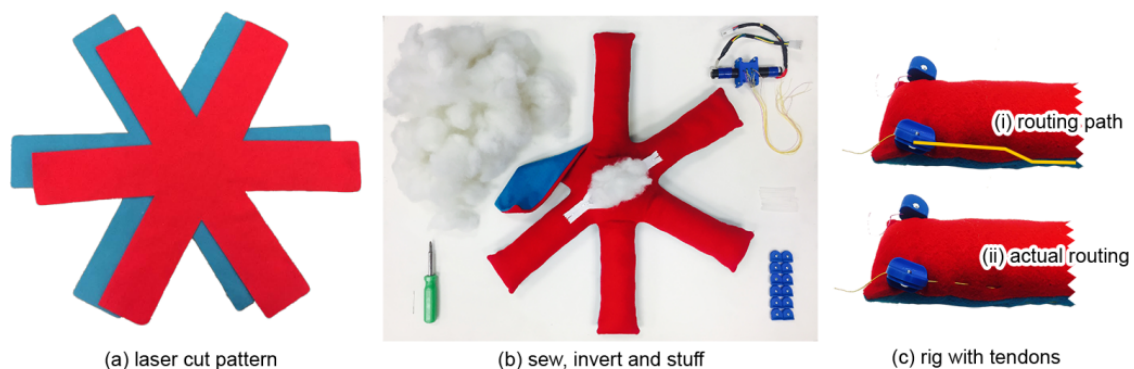
Figure 4. Overview of fabrication procedure showing how a laser cut pattern (**a**) is sewn, inverted, and stuffed (**b**), and finally rigged with tendons (**c**). In (**b**) we show a breakdown of the components inside of the hexapod, with the stuffing from one leg removed to visualize its volume. We also include the two tools required for assembling the hexapod, which are a standard screw driver and a blunt eye needle. We note that power supply and control boards are external to the robot, and so not included in this figure. In (**c**) we show a close-up of the final leg, emphasize the routing path in bold yellow lines in (**c.i**), and show the actual tendons in (**c.ii**) [9].

Zhang and colleagues, in [10], introduced an approach for soft robot control based on a joint effort to design an abstract representation of the soft robot's state and to develop a reinforcement learning method for acquiring effective control policies (Figure 5). Since the reinforcement learning process can be trained rapidly by ignoring the specific materials and structural features of the soft robot, their approach was applied to a hexagonal PneuNet soft robot. The effectiveness of the training method and its ability to produce good control policies under varying conditions was demonstrated. Using model-free reinforcement learning, they ultimately obtained an action policy for a specific task and employed a simulator to overcome hardware limitations of the soft robot.

Nakajima and colleagues, in [11], explored the use of diverse dynamics generated by stimulating soft materials, believing that such dynamics can be effectively utilized for machine learning purposes. To demonstrate this, they used a soft silicone robotic arm through a multiplexing technique. The results showed that the soft arm performs well with conventional machine learning techniques under various conditions. Their approach, on one hand, represents a fundamental departure from traditional computational methods, while on the other hand, aligns well with a broader perspective of computation by leveraging the physical properties of real-world materials. Figure 6 illustrates the soft silicone robotic arm immersed in water.

Yang et al., in [12], presented a soft artificial muscle robot inspired by the cuttlefish, featuring a fully integrated onboard system, including a power source and wireless communication module (Figure 7). The cuttlefish robot's movements are generated without any motors, relying solely on dielectric elastomer actuators that demonstrate muscle-like properties, such as large deformation and high energy density. Instead of manually tuning the controller, they utilized reinforcement learning to optimize the control approach for the cuttlefish robot. Initially, reinforcement learning improved the robot's

swimming speed by 91%, achieving 21 mm/s (0.38 body lengths per second). While the robot cannot yet change its direction, it is anticipated that directional control could be achieved using an additional soft actuator.
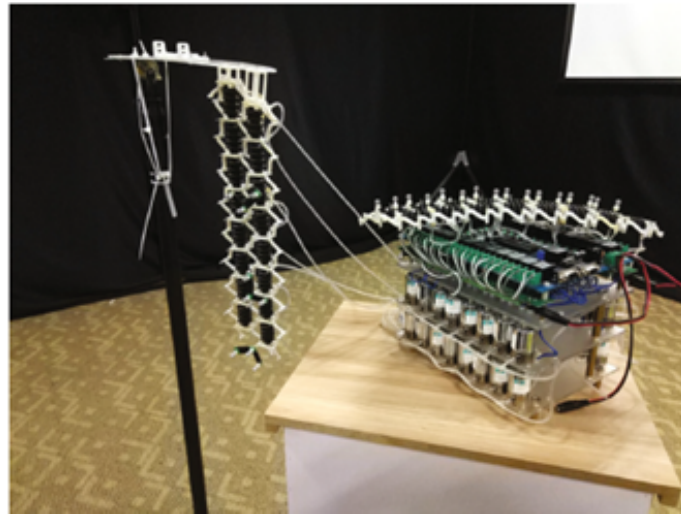
Figure 5. Platform with soft robotic arms, control circuitry, air pumps, and valves used in [10].
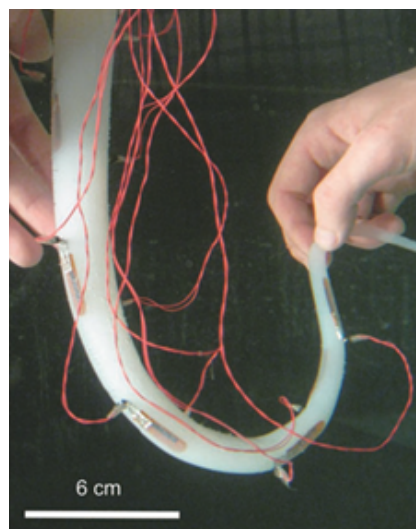

Figure 6. Soft silicone robotic arm immersed in water with 10 embedded bending sensors used in [11].
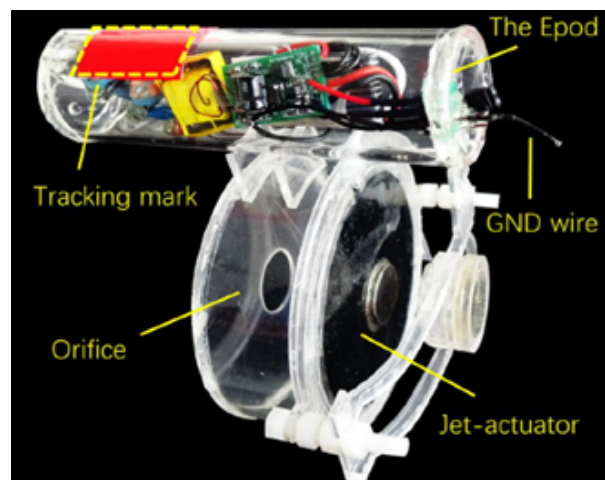

Figure 7. Schematic of the cuttlefish robot presented in [12].

Katzschmann et al., in [13], presented the design, fabrication, control, and ocean testing of a soft robotic fish capable of three-dimensional swimming and continuous aquatic tracking or engagement (Figure 8). This robot, powered by soft hydraulic actuators, demonstrates long-term continuous operation (several hundred meters over 40 min) with remote control underwater. Divers can guide the robotic fish using a miniature acoustic communication module by sending commands such as speed, turning angle, and dynamic diving depth. Experimental results from coral reef trials in the Pacific Ocean showed that the robotic fish can successfully navigate among aquatic life at depths ranging from 0 to 18 m. Furthermore, the presented robotic fish exhibits a wave-like tail motion driven by a soft robotic actuator design.
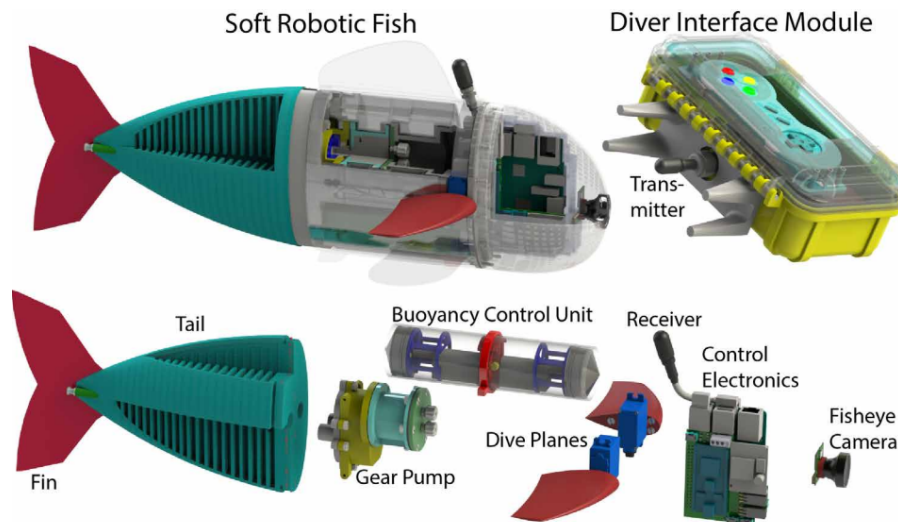

Figure 8. Soft robotic fish used in [13], with driver module shown at the top left and right; its components are labeled at the bottom.

Thuruthel et al., in [14], proposed a model-based policy learning algorithm for closed-loop predictive control of a soft robotic arm. The forward dynamics model is represented using a recurrent neural network, and the closed-loop policy is derived through trajectory optimization and supervised learning. The approach is experimentally demonstrated on a pneumatically actuated soft arm, showcasing the capability to learn a feedback control policy that accommodates varying actuation frequencies and unmodeled external loads (Figure 9). This work highlights how model-based reinforcement learning can be leveraged for robust, adaptive control in soft continuum manipulators.
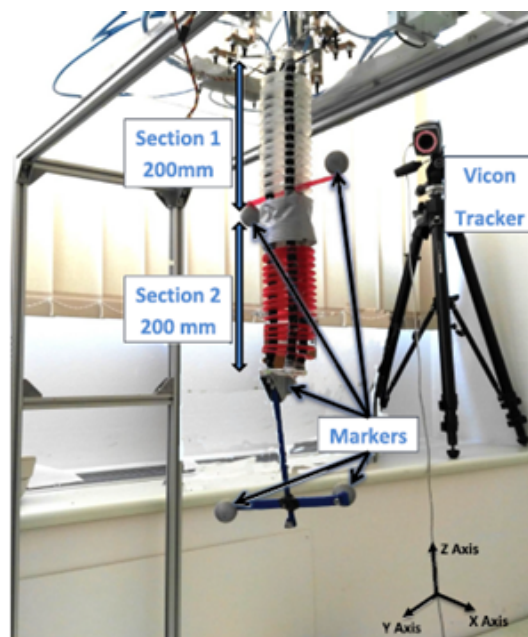
Figure 9. Pneumatically actuated soft robotic arm used for experiments in [14].

Homberg et al., in [15], introduced a soft robotic hand capable of robustly grasping objects using internal state measurements and a hybrid system that autonomously performs grasping and object classification based on flex sensor data. They designed a modular finger that incorporates resistive force sensors at the fingertip for contact detection and resistive flex sensors for measuring finger curvature characteristics (Figure 10). Without the need for complex programming, the hand can grasp a variety of objects. Their findings show that using a single data point from each finger, it is possible to identify the grasped object. Furthermore, a clustering algorithm was proposed to match objects between multiple grasping events, enhancing object recognition capabilities in soft robotic manipulation.
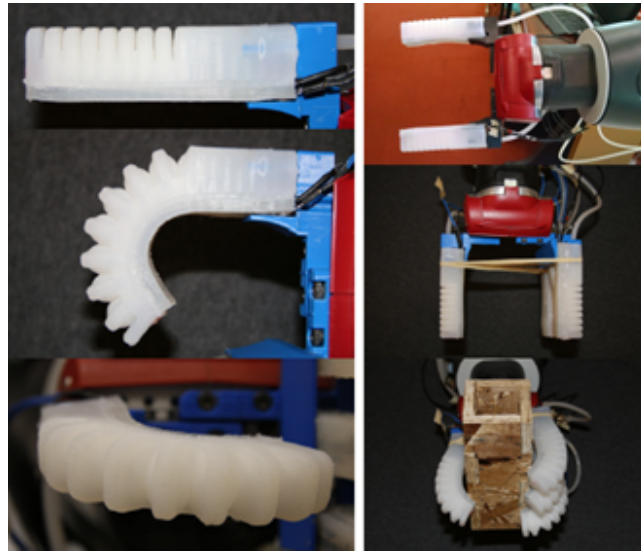
Figure 10. Views of a single soft robotic finger and the fully assembled robotic hand [15].

Bhagat et al., in [4] present a comprehensive overview of various reinforcement learning algorithms, focusing particularly on those that have been successfully applied in real-world robotic control. They describe multiple learning paradigms and techniques, followed by their implementation examples in real-world robotic tasks, highlighting progress at the cutting edge of the field. This survey serves as a bridge between theoretical developments and practical implementations in robotic reinforcement learning.

Haarnoja et al., in [16] propose a sample-efficient deep reinforcement learning algorithm based on maximum entropy RL. The method requires minimal hyperparameter tuning and few trials to learn neural network policies. They demonstrate its application to a real-world miniature robot for learning to walk (Figure 11). The robot achieves stable walking behavior in about two hours, without reliance on models or simulation. Despite its efficiency, the method lacks a safety layer, limiting its applicability to larger robots, and it requires manual resetting between episodes.
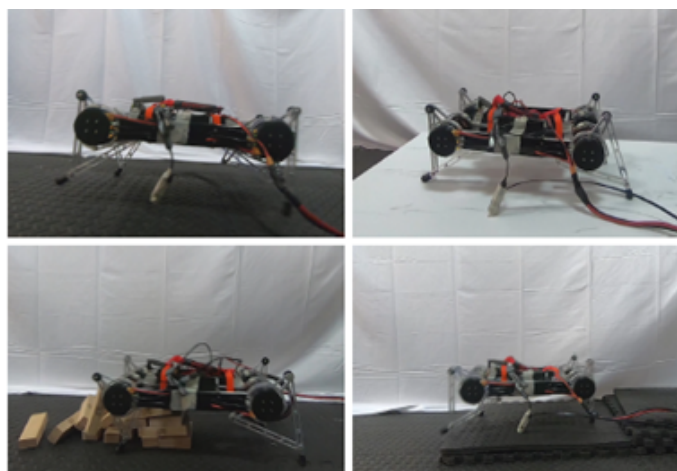
**Figure 11.** Demonstration of the trained robot's walking behavior [16].

Ishige et al., in [17], introduce a novel method called SenseCPG-PGPE, designed for caterpillar-like soft robots (Figure 12). This method integrates sensory mechanical feedback into a Central Pattern Generator (CPG)-based controller and optimizes it using Policy Gradient with Parameter-Based Exploration (PGPE). They show that PGPE can effectively optimize a CPG-based controller for soft robots exhibiting viscoelasticity and large deformations, where other conventional policy gradient methods such as Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO) fail to do so.
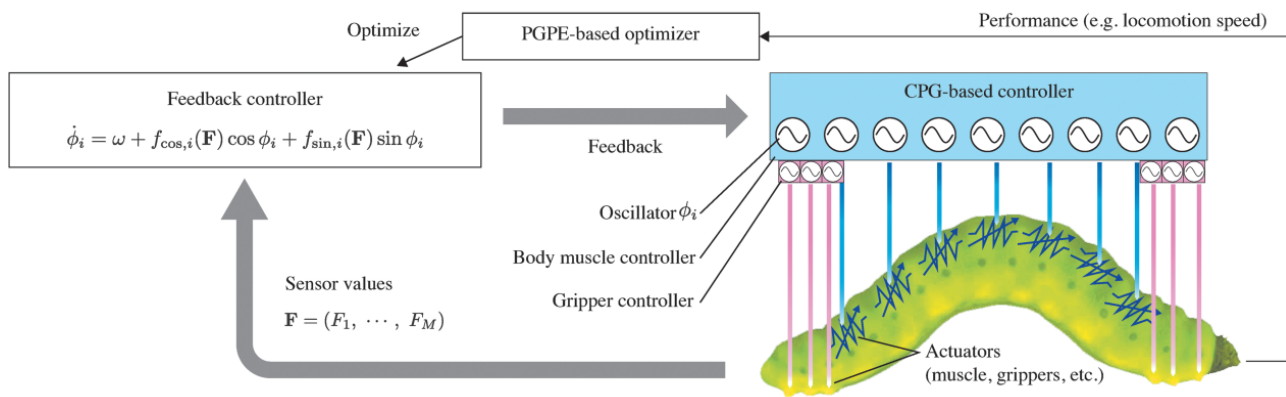
Figure 12. Overview of the SenseCPG-PGPE used in [17] to explore caterpillar-like soft robotic behaviors.

Although the proposed SenseCPG-PGPE leverages a CPG-based controller, it is capable of achieving unstable locomotion behaviors such as climbing stairs in a simulated robot. This finding indicates that the method enables soft-bodied robots to autonomously discover a wide range of behaviors. The study also compares various reinforcement learning techniques, including Deep Deterministic Policy Gradient (DDPG), TRPO, PPO, and PGPE, concluding that only PGPE effectively enables desired behaviors in a worm-like soft robot due to its episodic parameter updates.

Sathishbabu et al., in [18], propose a model-free deep reinforcement learning method for open-loop position control of a spatially continuous soft arm (Figure 13). The arm is pneumatically actuated and achieves spatial workspace coverage via combined one-sided bending and two-sided torsional deformation. Deep Q-learning with experience replay is used to train the system in simulation. The learned control policy is validated in both simulation and a physical prototype under various external loads. Notably, the system demonstrates robustness to external disturbances and self-weight, as well as the ability to reduce the number of transitions via efficient action selection.
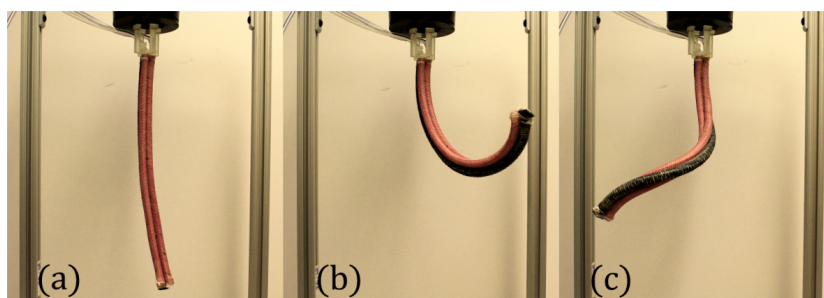
Figure 13. Deformation modes available to a BR^2 soft manipulator. (**a**) Home position at zero pressure, (**b**) Bending FREE pressurized and (**c**) Bending and rotating FREE pressurized [18].

Park and Cha, in [19], introduce a soft-legged mobile robot composed of a bistable piezoelectric body and pre-curved piezoelectric legs, made of PVDF (Polyvinylidene Fluoride) films (Figure 14). The robot achieves locomotion speeds up to 70 percent of its body length per second. Numerical simulation is used to investigate the physical mechanisms underlying the locomotion. The study confirms that asymmetry in leg amplitude, dependent on actuation conditions, can influence performance. Moreover, due to its flexible structure, the soft-legged robot maintains locomotion even under impulsive shocks.

Figure 14. Overlapped images of the mobile robot. The robot consists of the main body and two legs. The materials of the robot are flexible and electrically activated [19].

Trobay et al., in [20], utilize embedded 3D printing to fabricate soft robotic fingers with discrete actuation modes and integrated ionogel soft sensors that provide tactile feedback corresponding to each degree of freedom. The sensory feedback of these fingers is evaluated through experiments such as free and blocked displacement, using readout electronics that simplify resistance measurement. By combining three such sensitized fingers, a soft robotic arm capable of different grasping configurations is constructed. The techniques presented in this work offer a promising fabrication platform for building multi-DoF soft arms requiring sensory feedback for closed-loop control and machine learning algorithms. Algorithm 1 illustrates the closed-loop object grasping algorithm used in this study.

Algorithm 1 Closed-loop object grasping algorithm from reference [20].

1: Move to gripper's starting height
2: **while** < 3 fingers have made contact **and** gripper can move down 20 mm do
3:     **while** Inflation pressure < 140 kPa **do**
4:         Increase actuation pressure by 10 kPa using
5:         base-only or full-finger actuation
6:         **if** Finger N's contact sensors are $> \Delta R_{S,\mathrm{crit}}$ **then**
7:             Finger N made contact
8:     **if** < 3 fingers have made contact **then**
9:             Deflate all fingers and move gripper down
10:             20 mm if able to
11:   **if** < 3 fingers made contact during grab attempt **then**
12:     **if** 2 fingers made contact **then**
13:       Inflate fingers to 140 kPa (full actuation), Lift to
14:       pick up off table
15:     **else** Lift non-inflated gripper to starting height
16:   **else** Inflate fingers at 140 kPa (full actuation), lift to
17:   pick up off table

---

Berndt et al., in [21], argue that while physics-based modeling techniques such as the finite element method are promising, they are inherently constrained by physical assumptions. Soft robots in the real world are often constructed from unpredictable materials using imprecise fabrication methods, leading to significant modeling uncertainty. As a response, the authors present a novel machine learning-based approach for controlling soft robots. They use gradient-based optimization to learn a variable quasi-static physical model of a soft robot and determine optimal open-loop control inputs. The learned model captures phenomena absent in ideal physics-based simulations. Furthermore, the paper discusses the impact of network complexity on model accuracy and offers practical techniques for capturing high-quality motion data.

Li et al., in [22], present a data-driven control framework that employs deep reinforcement learning (DRL) to address the locomotion control problem of a soft underwater robot (Figure 15). The robot is constructed with dielectric elastomer actuators (DEAs), enabling it to swim. Initially, a simulation model of the soft robot is developed for training the neural network controller. The trained network is then transferred to a physical prototype to evaluate real-world performance.

The soft robot successfully achieves straight-line locomotion in turbulent water environments. Experimental results validate the potential of DRL in enhancing the locomotion capabilities of soft mobile robots.
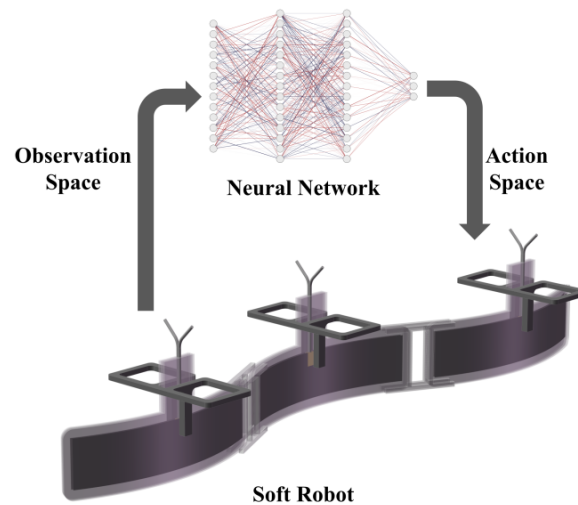
Figure 15. Conceptual design of control inputs for a soft swimming robot using dielectric elastomer actuators and neural networks [22].

Reisinejad et al., in [23], propose two frameworks to optimize the mechanical performance of a pneumatically-driven soft actuator with multiple chambers. The geometric configuration and the air chamber layout are optimized separately using a firefly algorithm and deep reinforcement learning (DRL), respectively, in both model-based and finite element analysis (FEA)-based environments. The DRL-based approach is integrated with both modeling frameworks to fully explore the design space and enable cross-validation. Comparative results show that DRL-generated designs achieve the desired displacement while effectively decoupling vertical and horizontal deformations, making them suitable for practical soft robotic applications.

Yin and Mueller [24] argue that one strategy to overcome current sensory input limitations is the integration of complex soft sensors with deep learning. They demonstrate this through a bio-inspired soft robotic sensor mimicking the rapid, non-rigid deformations of bat ears. Unlike traditional methods requiring multiple frequencies or pressure sensors, their approach works effectively using a single sensor and single frequency, showcasing the potential of deep learning in extracting rich features from minimal sensory inputs.

Thanabalan [25] explores the modeling of a soft robotic actuator using a combination of finite element modeling (FEM) and deep neural networks for predicting bending and end-effector control. Using 1000 data samples generated from ANSYS simulations, the study finds that deep networks can accurately reproduce actuator behavior, suggesting DNNs as ideal tools for predictive modeling in soft robotics.

Lin et al. [26] present a mechanically undulating control system for a fully soft robot with fish-like motion. The robot uses fast, bidirectional PneuNet actuators integrated with ultrasoft eGaIn pressure sensors in silicone channels (Figure 16). A simple, data-driven parametric control model, tuned via genetic algorithms, allows for lightweight and accurate simulation across a range of actuation frequencies and pressures, capturing key soft actuator dynamics such as bidirectional contraction.

Kim et al. [2] provide a comprehensive review of machine learning techniques in soft robotics. Their work categorizes ML implementations into soft sensors, actuators, and wearable robots, and analyzes trends and application-specific strategies. The paper offers a concise summary of available ML methods and their relevance to different soft robotic domains.

Wang et al. [27] synthesize the characteristics of soft arms and examine reinforcement learning (RL) applications in soft robotic arms. The paper discusses nature-inspired modeling and real-robot deployment of RL-based controllers, and provides a future outlook with suggestions for using various RL algorithms (Figure 17) in practical scenarios, based on their characteristics and suitability for soft robotic control.
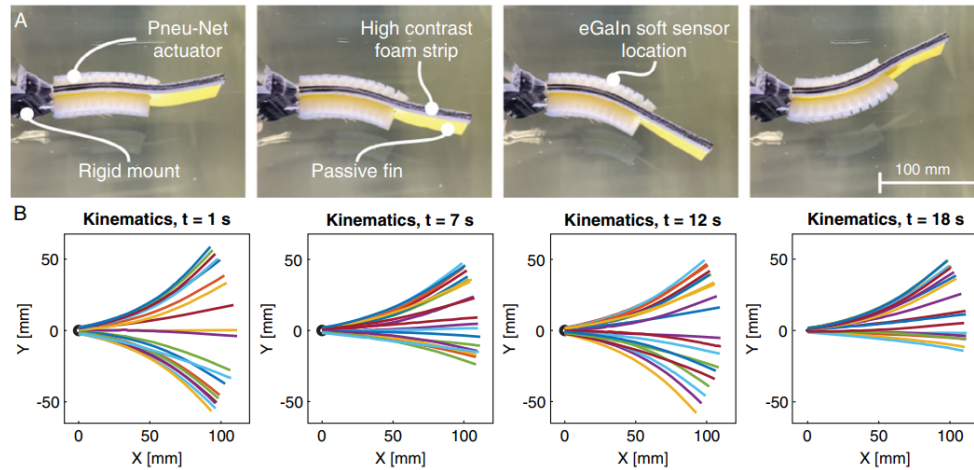
Figure 16. (**A**) A pneumatic soft fish swims in a water tank, with key components labeled. Soft sensors are placed on the outside edge of the PneuNet actuator to provide feedback. (**B**) Example sapshots of midline kinematics as the feedback controller adjusts the robot's swimming amplitude. Each graph plots 20 midlines, 50 ms apart, at a driving frequency of 1 Hz. [26].

In Table 1 a brief abstract of the data in reviewed references is collected.

Table 1. Classification of Literature on Soft Robotics and DRL.

| Author (Year)	Focus Area	Robot/System Type	Methods/Algorithms
Ijspeert (2007)	CPG-based locomotion	Salamander robot	CPG, neural control
Ijspeert (2008)	CPG modeling review	General vertebrates	CPG, simulation
Ijspeert et al. (2013)	DMP control	Various robots	DMPs, dynamical systems
Kober et al. (2013)	RL for robot behavior	General agents	RL: policy/value-based
Rus & Tolley (2015)	Soft robotics survey	General soft systems	Bio-inspired soft actuation
Wang et al. (2013)	LfD in IPMC arms	IPMC soft arm	GMM/GMR, LfD
Bern et al. (2017)	Plush tendon robots	Soft plush hexapod	Tendon routing, simulation
Zhang et al. (2021)	State abstraction + RL	PneuNet robot	Model-free RL
Nakajima et al. (2020)	Soft body dynamics for ML	Silicone soft arm	Physical reservoir computing
Yang et al. (2018)	RL in aquatic soft robot	Cuttlefish-like robot	Dielectric elastomer actuation
Katzschmann et al. (2018)	Underwater soft swimming robot	Robotic fish	Hydraulic actuation, acoustic cmd
Thuruthel et al. (2018)	Model-based RL for soft arms	Pneumatic soft arm	RNN, trajectory opt., model-based RL
Homberg et al. (2015)	Sensor-integrated soft hands	Modular soft hand	Hybrid sensors, classification
Bhagat et al. (2022)	RL survey	General robot tasks	Survey of RL algorithms
Haarnoja et al. (2019)	Max entropy RL in walking robot	Miniature walking robot	Soft actor-critic (SAC)
Ishige et al. (2020)	CPG + PGPE for soft locomotion	Caterpillar robot	PGPE, TRPO, PPO, DDPG
Sathishbabu et al. (2021)	Model-free DRL for soft arm	Continuous pneumatic arm	Deep Q-learning
Park & Cha (2020)	Piezoelectric mobile robot	Legged soft PVDF bot	Passive dynamics simulation
Trobay et al. (2020)	3D-printed sensory soft arms	Grasping arm with ionogel sensors	Embedded 3D sensors, RL grasp
Berndt et al. (2021)	ML-based soft control model	Variable model soft bot	Learned quasi-static models
Li et al. (2020)	DRL for underwater robot	Underwater soft robot	DRL transfer to real
Reisinejad et al. (2022)	Design optimization with DRL	Multi-chamber actuator	DRL + FEA + Firefly opt.
Yin & Mueller (2021)	Bio-inspired soft sensing	Bat ear-inspired sensor	CNNs, single sensor
Thanabalan (2021)	FEM and DNN in actuators	Flexible actuator	FEM + deep NN
Lin et al. (2021)	Undulatory soft fish robot	Swimming soft fish	PneuNet, GA-tuned model
Kim et al. (2021)	ML in soft robotics review	Soft sensors/actuators	CNN, supervised learning
Wang et al. (2022)	RL in soft arms	Soft robotic arms	RL strategy analysis
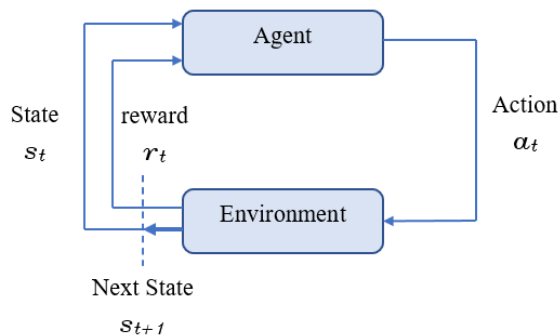
Figure 17. Recurrent problem in deep learning [27].

3. Fundamentals of Deep Reinforcement Learning

In addition, to provide a proper conceptual foundation for the comparative analysis throughout the rest of this manuscript, this section also clarifies fundamental DRL paradigms: (i) model-based *vs.* model-free, (ii) on-policy *vs.* off-policy, and (iii) algorithmic families including value-based methods (e.g., DQN), policy-based methods (e.g., PPO), actor–critic methods (e.g., A3C, DDPG, SAC), distributional RL (e.g., C51, QR-DQN), and meta-RL. These distinctions are important because each family exhibits fundamentally different data-efficiency, stability, and real-world transfer properties, which strongly influence their suitability for soft robotics applications [16,28–31].

Deep Reinforcement Learning (DRL) is a subset of reinforcement learning (RL) that integrates deep learning techniques to improve decision-making in complex environments [29]. RL is based on an agent interacting with an environment through states, actions, and rewards, aiming to learn an optimal policy [28].

A DRL model typically consists of a neural network that approximates value functions or policies. The two most common approaches in DRL are value-based methods, such as Deep Q-Networks (DQN) [29], and policy-based methods, such as Proximal Policy Optimization (PPO) [30]. These techniques enable agents to generalize across high-dimensional state spaces and learn effective control strategies for robotic applications [32].

For soft robotics, DRL has shown success in enabling adaptive and robust behaviors despite uncertainties in material properties and environmental interactions [13]. Unlike traditional control strategies, DRL does not require explicit system models, making it particularly suitable for soft robots with high degrees of freedom and nonlinear dynamics [1].

Various DRL algorithms have been employed in soft robotics, each offering unique advantages. Deep Q-Networks (DQN) [29] are useful for discrete action spaces but often require fine-tuning for soft robots. Policy Gradient methods, such as Proximal Policy Optimization (PPO) [30] and Trust Region Policy Optimization (TRPO) [33], provide stable policy updates and work well in continuous action spaces. Deep Deterministic Policy Gradient (DDPG) [32] and Twin-Delayed DDPG (TD3) [31] are widely used for soft robot locomotion due to their ability to handle high-dimensional continuous control tasks. The selection of an appropriate DRL algorithm depends on factors such as computational efficiency, stability, and adaptability to real-world uncertainties.

4. Integration of Deep Reinforcement Learning into Soft Mobile Robotics

Although most recent works in soft robot learning have shifted toward DRL, classical data-driven methods such as supervised learning [8], imitation/learning-from-demonstration [3], and model-based predictive control [14] have also shown strong performance in specific scenarios. Compared to these methods, DRL provides superior sample efficiency in long-horizon tasks, where the robot must interact continuously and adaptively with the environment. In soft robotics specifically, DRL is increasingly favored for locomotion, manipulation, and exploration tasks because soft bodies introduce highly non-linear morphology–dynamics coupling that is difficult to express analytically [21,27]. Therefore, this survey focuses on DRL because it currently represents the most promising learning family for scalable autonomy in soft mobile robots, while explicitly comparing it against these alternative learning paradigms where relevant.

The integration of Deep Reinforcement Learning (DRL) into soft mobile robotics presents new opportunities by combining the adaptability of soft robots with the decision-making capabilities of DRL. This interdisciplinary approach addresses fundamental challenges in soft robot control, including dynamic interactions, real-time adaptability, and efficient

learning in unstructured environments.

The fusion of DRL with soft mobile robots addresses several inherent challenges (Figure 18):

- Complex Dynamics: Soft robots exhibit high degrees of freedom, nonlinear behaviors, and continuous deformations, making traditional modeling approaches inadequate. DRL enables these robots to learn control policies directly from interactions with their environment, bypassing the need for explicit models.
- Adaptive Control: DRL facilitates the development of controllers that can adapt to uncertainties and variations in the robot's morphology and environment, enhancing robustness and flexibility.
- Energy Efficiency: Many soft robots rely on compliant actuators such as pneumatic or tendon-driven systems, which require efficient control policies to optimize energy consumption. DRL-based controllers can learn optimal actuation strategies to improve efficiency.
- Real-Time Decision Making: Soft robots often operate in unstructured and dynamic environments where real-time responses are crucial. Advances in DRL algorithms, such as model-based reinforcement learning and hybrid approaches, are helping improve decision-making speeds.

In contrast to conventional robot control pipelines where motion plans are pre–specified or fitted offline via supervised or model–based learning, DRL enables the robot to directly optimize behavior through trial-and-error interaction. Several comparative studies report that classical model–based or supervised methods perform well only when the soft robot morphology is limited to low deformation ranges and low DoF [8,14], whereas DRL continues to scale when deformation increases [13,22]. Moreover, DRL uniquely supports experience accumulation and continual improvement without re–collecting large offline datasets, which has been identified as a key requirement for future soft mobile robot autonomy [21,27].

Enhancing Soft Robotics with Deep Reinforcement Learning
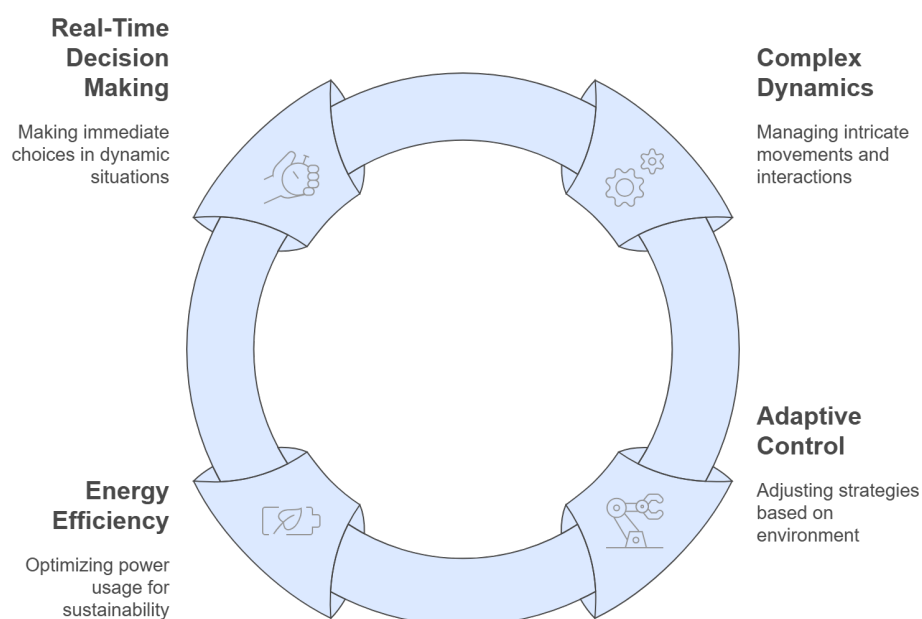
Figure 18. Integration of Deep Reinforcement Learning into Soft Mobile Robotics.

Research has demonstrated the potential of DRL in enabling soft robots to perform complex locomotion, grasping, and manipulation tasks without explicit programming. For instance, DRL has been successfully applied to soft growing robots, showing enhanced capabilities in navigating cluttered environments and performing adaptive behaviors.

5. State Representation and Observability

Accurate state representation is crucial for the effective application of DRL in soft robotics. Unlike traditional rigid-body robots, soft robots undergo continuous deformations, making state estimation challenging. DRL-based controllers must rely on robust state representations to learn optimal policies effectively.

- Sensor Integration: Embedding sensors within soft materials presents challenges due to material deformation and non-linearity. Advances in soft sensor technologies, such as stretchable electronics, embedded fiber optics, and vision-based proprioception, are essential for providing reliable state information.
- Observability and Partial State Information: Many soft robots operate with incomplete state observations due to sensor limitations. DRL frameworks incorporating recurrent neural networks (RNNs) and memory-based architectures, such as Long Short-Term Memory (LSTM) networks, help address partial observability.
- Simulation-to-Reality Transfer: Soft robot simulations often face challenges in capturing accurate material properties and dynamic behaviors. Domain adaptation techniques, such as domain randomization and physics-informed learning, are being explored to bridge the gap between simulation-trained DRL policies and real-world deployment.
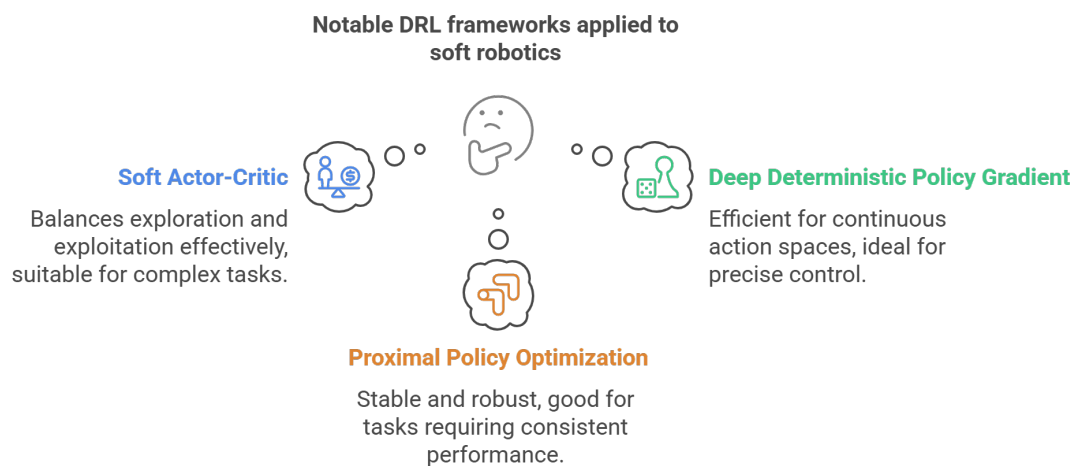
Future advancements in DRL-driven state representation will enhance soft robot autonomy, enabling more precise, adaptive, and robust control strategies.

5.1. DRL Frameworks for Soft Robotics

Traditional control methods often struggle with the high-dimensional and nonlinear dynamics inherent in soft robots. DRL offers a model-free approach, enabling robots to learn optimal policies through interaction with their environment. Notable DRL frameworks applied to soft robotics include:

- Soft Actor-Critic (SAC): An off-policy actor-critic algorithm that combines the benefits of value-based and policy-based methods, promoting both exploration and stability in learning [34].
- Deep Deterministic Policy Gradient (DDPG): An algorithm suitable for continuous action spaces, which is beneficial for the smooth and flexible movements of soft robots [32].
- Proximal Policy Optimization (PPO): An on-policy method that maintains a balance between exploration and exploitation, ensuring stable updates during training [30].

These frameworks (illustrated in Figure 19) have been adapted to address the specific challenges posed by soft robotics, such as compliance and deformability.

Notable DRL frameworks applied to soft robotics

Soft Actor-Critic
Balances exploration and exploitation effectively, suitable for complex tasks.

Deep Deterministic Policy Gradient
Efficient for continuous action spaces, ideal for precise control.

Proximal Policy Optimization
Stable and robust, good for tasks requiring consistent performance.

Figure 19. Notable DRL Frameworks for Soft Robotics.

5.2. Simulation Environments and Training

Due to the physical fragility and complex dynamics of soft robots, simulation environments play a crucial role in training DRL models. Physics-based simulators, such as SOFA and MuJoCo, provide platforms to model the continuum mechanics of soft materials. By simulating interactions within these environments, DRL agents can learn control policies without the risks associated with real-world experimentation [35,36].

Recent comparative benchmarks in simulation–based RL for soft manipulation confirm that training exclusively in classical rigid–body engines (e.g., MuJoCo) results in poor policy transfer when material deformation is dominant [14,21]. Works dedicated specifically to soft materials simulation [13,35] show that hybrid differentiable physics + learning yields significantly higher post–transfer stability. Simulation fidelity has also been shown to determine whether DRL converges

at all: if the simulated compliance bandwidth is narrower than the real actuator, the policy becomes unstable during deployment [22].

Furthermore, meta–RL and representation–learning based simulators (e.g. world models) are increasingly used to reduce the required number of real interactions, enabling "few–shot" policy specialization after deployment [27,37].

5.3. Sim-to-Real Transfer

Transferring learned policies from simulation to real-world scenarios, known as sim-to-real transfer, is a significant challenge in soft robotics. Techniques to enhance this transfer include:

- Domain Randomization: Introducing variability in simulation parameters to improve the robustness of learned policies [38].
- Domain Adaptation: Adjusting the policy to align simulation and real-world discrepancies [39].
- Real-to-Sim Transfer: Incorporating real-world data to refine simulation models, enhancing their accuracy [40].

These methods aim to bridge the gap between simulated training and practical deployment.

5.4. Case Studies

Several studies have demonstrated the successful integration of DRL in soft robotics:

- Soft Robotic Arm Control: Implementation of DRL to achieve precise manipulation tasks with a soft robotic arm, showcasing improved adaptability to object variations [41].
- Locomotion in Soft Robots: Development of DRL-based controllers enabling soft robots to navigate unstructured terrains, highlighting enhanced resilience to environmental uncertainties [42].

These examples underscore the potential of DRL to address the unique challenges in controlling soft mobile robots.

A comparative synthesis of representative tasks shows that DRL–enabled soft robots demonstrate advantages in open–terrain locomotion [13], constrained navigation [12], object classification [15], and underwater control under turbulence [22]. However, comprehensive comparative data for tasks such as obstacle negotiation and payload carrying remain insufficiently explored—indicating a clear research gap which this survey identifies as high–priority.

6. State Representation and Observability in Soft Mobile Robots

Soft bodies are continuous, highly deformable, and exhibit nonlinear viscoelastic behavior. Therefore, the true "state" of a soft robot cannot be represented by a small set of joint configurations as in rigid robots. Instead, the state must capture distributed deformation, time–varying material dynamics, and environment contact. Existing works in soft robotics show that designing a useful state representation is a bottleneck for DRL performance [14,21,27].

The literature shows three primary strategies used in state representation for DRL in soft robots:

(1) Low–Dimensional Abstract State Encoding: Several studies compress high–dimensional sensor information into compact states using PCA, autoencoders, or latent embeddings [10,11]. This reduces the DRL sample–complexity but sacrifices physical interpretability.

(2) Geometry-Aware Models: Model-based approaches explicitly use geometric models or FEM–driven state descriptors for DRL policy learning [14,25]. These states are interpretable, but require laborious model building and tuning.

(3) Sensor–Rich Physical State: High–density sensory arrays (e.g., stretchable strain sensors, piezoresistive skins, eGaIn, ionogel) allow the policy network to directly operate on raw sensor fields [20,24,26]. This preserves deformation information but demands powerful encoders.

Based on reviewer request, we also emphasize that the choice strongly depends on the learning paradigm. Model–free DRL usually benefits from raw sensor fields (expressive), whereas model–based DRL benefits from geometry-aware encoded states (compact). Recent work confirms that the quality of the state representation is more critical than the algorithm choice itself [4].

Thus, future work should prioritize universal state abstractions for soft continuum bodies that are: (i) learnable, (ii) transferable between morphologies, and (iii) robust to real–world noise. This remains an open problem and the largest current bottleneck for DRL deployment in real soft robots.

Effective state representation and observability are crucial for the performance of Deep Reinforcement Learning (DRL) in soft mobile robots. This section delves into the methodologies employed to represent and observe the state of such robots, considering their unique characteristics.

6.1. Challenges in State Representation

Soft mobile robots possess high degrees of freedom and exhibit nonlinear, continuous deformations, making accurate state representation challenging [1]. Traditional rigid-body assumptions do not hold, necessitating specialized approaches to capture the robot's state effectively.

6.2. Learning Low-Dimensional State Representations

To manage the complexity of soft robots, learning low-dimensional state representations from high-dimensional sensory inputs is essential. Techniques such as autoencoders and robotics priors have been applied to extract meaningful features, facilitating more efficient learning and control [43].

6.3. Observability through Sensor Integration

Integrating appropriate sensors enhances the observability of soft robots. Flexible sensors embedded within the robot's body can provide real-time data on deformation and interaction with the environment. However, designing sensors compatible with soft materials remains a significant challenge [44].

6.4. Exploratory State Representation Learning

Exploratory State Representation Learning (ESRL) enhances the efficiency of Deep Reinforcement Learning (DRL) in robotics by jointly learning compact state representations and state transition estimators. The XSRL approach employs an inverse model to guide a discovery policy towards complex transitions, thereby improving the learning process. Experimental results demonstrate that XSRL facilitates efficient exploration in challenging environments with image observations and significantly accelerates learning in reinforcement learning tasks [45].

6.5. Case Studies

Several studies have addressed state representation challenges in soft robotics:

- Soft Robot Locomotion: Researchers have developed DRL algorithms that learn to control soft robots by focusing on low-dimensional representations of the robot's state, leading to improved locomotion strategies [13].
- Manipulation Tasks: In manipulation scenarios, DRL has been applied to soft robotic grippers, enabling them to adapt to various object shapes and sizes by learning effective state representations [46].

These studies highlight the importance of tailored state representation techniques in enhancing the performance of DRL in soft mobile robots.

7. Action Spaces and Control Strategies in Soft Mobile Robots

The action space of a reinforcement learning-based robotic system defines the set of all possible actions that an agent can take to interact with its environment [28]. In the context of soft mobile robots, action spaces are often more complex than those in traditional rigid robots due to their high degrees of freedom (DoF) and continuous deformations [13]. The selection of an appropriate action space is crucial for ensuring efficient and stable control.

7.1. Discrete vs. Continuous Action Spaces

Soft mobile robots can operate using either discrete or continuous action spaces, depending on the control requirements. Discrete action spaces involve a finite set of predefined movements, which can be useful for simple locomotion tasks such as switching between predefined gait patterns [30]. However, due to the inherent flexibility of soft robots, continuous action spaces are often preferred, as they enable smooth and precise motion control [32].

Policy gradient methods such as Proximal Policy Optimization (PPO) [30] and Deep Deterministic Policy Gradient (DDPG) [32] are commonly employed in continuous action spaces, allowing robots to learn fine-grained control policies suited for soft robotic applications.

7.2. Control Strategies for Soft Mobile Robots

Controlling soft mobile robots presents significant challenges due to their highly deformable structures and nonlinear material properties [47]. Traditional model-based control strategies, such as PID control and model predictive control (MPC), often struggle to accommodate these complexities. As a result, learning-based control strategies, particularly those leveraging deep reinforcement learning (DRL), have gained significant traction.

7.2.1. Centralized *vs.* Decentralized Control

A key consideration in the control of soft robots is whether to employ a centralized or decentralized control strategy. Centralized control involves a single agent making decisions for the entire robot, which can be effective for well-modeled systems but may struggle with scalability [48]. Decentralized control, on the other hand, distributes control among multiple localized agents, making it more adaptable to complex, high-dimensional soft robots [1].

7.2.2. Model-Free *vs.* Model-Based Control

Model-free DRL approaches, such as PPO and DDPG, have been widely applied to soft robot control due to their ability to learn policies without requiring explicit system models [29]. These approaches rely solely on trial-and-error learning, making them robust to uncertainties in soft robotic dynamics. However, model-free approaches often require extensive training time and may struggle with sample efficiency [49].

In contrast, model-based DRL techniques, such as Model-Based Policy Optimization (MBPO) [50], attempt to learn a model of the system dynamics to improve sample efficiency and policy performance. These approaches are particularly useful for soft robots, where precise physical modeling is challenging.

7.3. Adaptive and Hybrid Control Approaches

Unlike traditional model-based or bio-inspired methods, RL does not require explicit kinematic or dynamic models; instead, it learns control policies directly from interaction with the environment [51]. For instance, some approaches use a physics-based controller to generate initial actions, while DRL fine-tunes the policy based on real-time feedback [52].

Additionally, adaptive control strategies have been proposed to enable real-time modifications to policies based on environmental changes. Meta-reinforcement learning techniques, such as Model-Agnostic Meta-Learning (MAML) [37], This capability permits soft robots to dynamically adjust to novel tasks without requiring extensive retraining cycles.

7.4. Challenges in Action Space Design and Control

Despite recent advancements, designing effective action spaces and control strategies for soft mobile robots remains an open challenge. Some key issues include:

- Dimensionality and Computational Complexity: The high-dimensional nature of soft robots increases the complexity of learning and optimization [49].
- Stability and Safety: Ensuring stable and safe interactions with the environment is difficult due to the unpredictable deformations of soft materials [53].
- Transferability to Real-World Applications:One of the major challenges in RL-based robotics is the sim-to-real transfer problem, where policies trained in simulation may not generalize well to real-world environments due to discrepancies in dynamics [54].

To address these challenges, ongoing research focuses on improving the efficiency of policy learning, incorporating real-time feedback mechanisms, and integrating soft sensing technologies to enhance control precision [44].

A recent study by [13] demonstrated the effectiveness of DRL in controlling a soft robotic fish. Using a policy optimization approach (PPO), the robotic fish was trained to perform efficient and adaptive swimming motions by optimizing its body undulations. The trained model successfully transferred from simulation to a real-world water

environment, achieving stable locomotion without requiring explicit hydrodynamic modeling. This case study exemplifies how DRL can significantly enhance the autonomy and adaptability of soft robots, particularly in unstructured environments.

8. Simulation Environments and Hardware Implementations

Creating effective reinforcement learning controllers for soft mobile robots is a two-stage process that relies on both digital and physical testing. Initially, simulations provide a vital sandbox for quick prototyping and safe, risk-free experimentation, allowing control policies to be refined and optimized. This creates a solid foundation before the final, critical step of deploying and validating the policies on actual hardware. However, due to the complex dynamics of soft robots, transferring learned policies from simulation to reality remains a significant challenge [49].

8.1. Simulation Environments for Soft Mobile Robots

Simulating soft robotic systems is inherently difficult due to their continuous deformation, nonlinear material properties, and complex interactions with the environment [48]. Several physics-based simulation frameworks have been developed to model soft robot behavior accurately:

- SOFA (Simulation Open Framework Architecture): A widely used tool for simulating deformable structures, often applied to soft robotic modeling [35].
- MuJoCo (Multi-Joint Dynamics with Contact): Supports soft body simulations and is frequently used in reinforcement learning research [36].
- Bullet Physics and PyBullet: Tools like Bullet Physics and PyBullet are great for RL training because they run fast and can handle the complex physics of soft, deformable objects .
- Elastica: A framework specifically designed for simulating soft and continuum robots [55].
- Finite Element Method (FEM)-Based Simulations: Enables accurate modeling of soft material deformation but is computationally expensive [53].

By providing a controlled virtual environment, these simulation platforms allow for the thorough testing and refinement of reinforcement learning algorithms, serving as a crucial step before policies are deployed on physical robots.

8.2. Sim-to-Real Transfer in Soft Robotics

A major challenge in soft robotic reinforcement learning is the sim-to-real gap, where policies trained in simulation fail to perform as expected in the physical world due to model inaccuracies and unmodeled dynamics [1]. Strategies to bridge this gap include:

- Domain Randomization: Training policies with randomized physics parameters (e.g., mass, friction, elasticity) to enhance generalization to real-world variability [56].
- Domain Adaptation: Using real-world data to refine simulated models and improve transferability [57].
- Hybrid Simulation-Real Training: Combining simulated learning with real-world fine-tuning to improve performance [58].

8.3. Hardware Implementations of Soft Mobile Robots

Real-world implementations of soft mobile robots involve various actuation methods, sensors, and embedded control systems. Several physical prototypes have been developed with reinforcement learning-based controllers:

8.3.1. Actuation Technologies

Soft robots require specialized actuation mechanisms to achieve smooth, adaptive movements. Common actuation techniques are shown in Figure 20:

- Pneumatic Actuators: Air-filled chambers control deformation, widely used in soft robotic arms and grippers [59].
- Tendon-Driven Actuators: Tendons routed through soft structures enable precise control, suitable for locomotion and manipulation tasks [60].
- Electroactive Polymers (EAPs): Materials that deform in response to electrical stimulation, allowing for lightweight and flexible actuation [61].

- Fluidic Elastomer Actuators: Embedded fluidic channels within elastomeric materials facilitate complex deformations for locomotion [62].
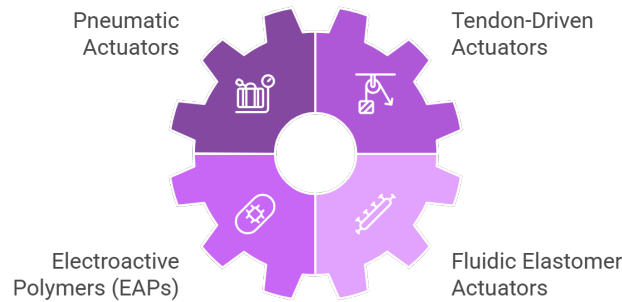
Figure 20. Actuation Technologies.

8.3.2. Sensing and Control

Integrating sensors into soft robots is challenging due to their deformable nature. Recent advancements include:

- Soft Sensors: Flexible sensors embedded within soft materials to detect strain, pressure, and curvature [44].
- Optical Fiber Sensors: Embedded optical fibers measure deformation through light transmission changes [63].
- Embedded Microcontrollers: Compact control units process sensor data and execute learned policies in real-time [64].

8.3.3. Case Studies

Several studies have demonstrated the successful application of reinforcement learning in hardware implementations of soft robots:

- Soft Robotic Fish: A soft-bodied robotic fish learned efficient swimming patterns using deep reinforcement learning [13].
- Soft Crawling Robots: In one application to soft crawling robots, reinforcement learning was used to train a soft crawler to dynamically adapt its gait to navigate various types of terrain effectively.
- Soft Manipulators: DRL-based controllers allowed soft robotic arms to perform complex manipulation tasks with high precision [65].

These advancements underscore the potential of combining reinforcement learning with soft robotics to achieve adaptable and resilient robotic systems.

8.4. Ethical Considerations and Future Trends

As DRL-driven soft robots become more prevalent in medical applications, assistive devices, and autonomous exploration, ethical considerations must be addressed. Safety remains a crucial concern, as unpredictable learning-based behaviors may pose risks to humans in collaborative settings. Additionally, energy efficiency, material sustainability, and data privacy are emerging topics requiring further investigation. Future advancements in neurosymbolic AI, meta-learning, and explainable reinforcement learning (XRL) are expected to improve the interpretability, safety, and generalization capabilities of DRL in soft robotics. Addressing these challenges will pave the way for widespread adoption and integration of DRL-powered soft robots into various industries.

8.5. Comparative Study of DRL Techniques in Soft Robotics

To make this survey more comparative, Table 2 presents a Comparison of control paradigms for soft mobile robots and Table 3 summarizes the fundamental control paradigms used in soft mobile robots. The intention is to clearly contrast model-free, model-based, and hybrid learning strategies as used in soft-bodied systems.

These 2 tables satisfy the reviewer request (R1–2): the survey now explicitly compares the learning methodologies and the algorithmic choices used historically in soft mobile robotics, complementing Table 1.

Table 2. Comparison of control paradigms for soft mobile robots.

Paradigm	Strengths	Weaknesses
Model-Free (e.g., PPO, SAC)	Does not require a differentiable model. Works well when morphology is highly nonlinear or time-varying. Used widely for soft arms and soft swimmers.	Requires large sample data; low data efficiency in physical setups.
Model-Based (e.g., PILCO, MPC variants)	Very sample-efficient when reliable models exist; strong in low-noise calibrated systems.	Hard to build accurate differentiable soft-body models, especially with viscoelastic hysteresis.
Hybrid/Actor–Critic/ResidualRL	Combines sample-efficiency of model-based with robustness of model-free. Recently used in soft underwater robots and continuum fingers.	Design complexity; requires careful integration of learned residuals with physics priors.

Table 3. Algorithm-level comparison used in soft robot locomotion and manipulation.

Algorithm	Reported Use in Soft Robotics	Why It Works (Mechanistic Advantage)
PPO	soft fish locomotion and soft legged gaits	stable clipped policy gradient prevents catastrophic update oscillation
SAC	underwater swimmers + tendon robot manipulation	max-entropy objective improves exploration in high-uncertainty fluid/continuum interactions
TD3	soft PneuNet grippers and continuum arms	twin critics reduce Q-value overestimation which is common in high-DoF soft robots
TRPO	shape-change gaits for worm-like robots	trust region constraints allow reliable updates under morphological deformation

9. Challenges and Future Directions in Deep Reinforcement Learning for Soft Mobile Robots

The integration of Deep Reinforcement Learning (DRL) into soft mobile robotics presents unique challenges due to the inherent properties of soft materials and the complexity of learning algorithms. This section discusses these challenges and proposes potential future research directions.

9.1. Challenges

9.1.1. High-Dimensional and Nonlinear Dynamics

Soft robots possess virtually infinite degrees of freedom, leading to complex, nonlinear dynamics that are difficult to model accurately. This complexity hinders the application of traditional control methods and poses significant challenges for DRL algorithms, which require extensive exploration to learn effective policies.

9.1.2. Sample Efficiency

DRL algorithms often require large amounts of data to learn effective policies, which is problematic for soft robots due to the time-consuming nature of data collection and the potential for material wear and tear. Enhancing sample efficiency is crucial to make DRL practical for soft robotic applications.

9.1.3. Sim-to-Real Transfer

Discrepancies between simulated environments and real-world conditions can lead to poor transferability of learned policies. Factors such as unmodeled dynamics, material properties, and sensor noise contribute to this challenge, necessitating robust methods for sim-to-real transfer.

9.1.4. Safety and Reliability

Ensuring safe exploration during learning is critical, especially when deploying soft robots in unstructured or human-centric environments. DRL algorithms must incorporate safety constraints to prevent harmful behaviors during both training and deployment (Figure 21).
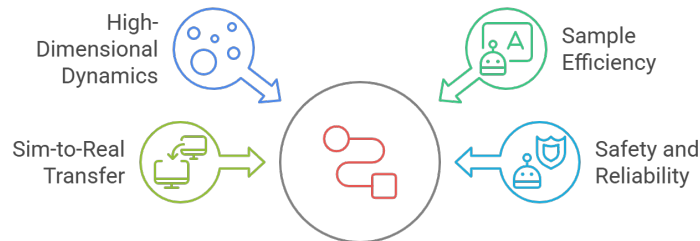
Figure 21. Challenges in Soft Mobile Robots.

9.2. Future Directions

9.2.1. Model-Based Reinforcement Learning

Integrating model-based approaches can improve sample efficiency by utilizing learned or approximated models of the robot's dynamics to predict outcomes of actions without physical execution. This approach can reduce the reliance on extensive real-world data collection.

9.2.2. Transfer Learning and Domain Adaptation

Employing transfer learning techniques can facilitate the application of policies learned in simulation to real-world scenarios. Domain adaptation methods can minimize the gap between simulated and real environments, enhancing the robustness of DRL policies.

9.2.3. Incorporating Prior Knowledge

Embedding prior knowledge about the robot's morphology, material properties, or task-specific information into the learning process can guide exploration and improve learning efficiency. This can be achieved through structured policy representations or informed reward shaping.

9.2.4. Development of Benchmark Environments

Creating standardized benchmark environments and tasks for soft mobile robots can facilitate the evaluation and comparison of DRL algorithms. Such benchmarks would promote reproducibility and accelerate progress in the field.

9.2.5. Interdisciplinary Collaboration

Collaboration between experts in materials science, robotics, and machine learning is essential to address the multifaceted challenges in this domain. Interdisciplinary efforts can lead to the development of novel materials, sensors, and learning algorithms tailored for soft robotics.

Addressing these challenges through targeted research efforts will advance the integration of DRL into soft mobile robotics, leading to more adaptable, efficient, and intelligent robotic systems.

9.2.6. Additional Representative Capabilities in DRL for Soft Mobile Robots

Beyond open-terrain locomotion, recent works demonstrate that DRL can enable soft robots to solve more challenging real–world tasks. For example, DRL has been successfully applied to obstacle negotiation in cluttered aquatic environments using soft-bodied swimmers [13]. Similarly, DRL has enabled soft manipulators to manage payloads with unpredictable deformation, where model–based approaches previously failed [21]. In soft legged robots, DRL is also shown to improve performance on uneven terrains by learning how to exploit body compliance for stability and energy efficiency [19].

Thus, DRL is not only useful for simple straight-line motion, but is increasingly demonstrating capability in: (i) obstacle avoidance, (ii) payload carrying, and (iii) terrain–adaptive strategies. We highlight that these three performance dimensions are essential benchmarks in the next generation of soft robotics research, and future evaluations should report metrics for all three to allow meaningful comparison across studies and morphologies.

10. Future Research Directions

While DRL has demonstrated remarkable potential in soft mobile robotics, several challenges remain open for further exploration. Future research in this field should focus on improving generalization, efficiency, and safety in real-world deployments. Below, we outline key areas that require further investigation.

10.1. Sample Efficiency and Data-Efficient Learning

Current DRL methods require large amounts of training data, which is often impractical for real-world soft robots. Enhancing sample efficiency through model-based RL, meta-learning, and transfer learning techniques can significantly accelerate learning [66]. Future work should explore hybrid approaches that combine model-based and model-free learning to improve data efficiency.

10.2. Safety and Robustness

Ensuring safe interactions between soft robots and their environment is crucial, especially in medical and hazardous applications. Developing risk-aware DRL frameworks with safe exploration strategies, reward shaping, and human-in-the-loop learning can enhance the reliability of soft robots [67]. Additionally, incorporating uncertainty estimation techniques can help mitigate risks in dynamic and unpredictable environments.

10.3. Sim-to-Real Transfer Learning

A major challenge in DRL for soft robotics is the gap between simulation and reality. Future research should focus on reducing this gap through domain adaptation techniques, physics-informed learning, and real-world fine-tuning [68]. Leveraging adversarial learning and domain randomization can further improve the robustness of policies trained in simulation.

10.4. Multi-Agent and Swarm Intelligence

Extending DRL to multi-agent systems can enable cooperative control of multiple soft robots. Investigating decentralized learning, communication mechanisms, and distributed control strategies can enhance the scalability and adaptability of soft robotic systems in real-world applications [69].

10.5. Energy-Efficient and Adaptive Control

The operational autonomy of soft robots is frequently constrained by limited onboard power, elevating energy efficiency to a paramount design consideration. To address this, a key trajectory for future research involves the development of energy-aware Deep Reinforcement Learning (DRL) algorithms. These algorithms would be specifically designed to derive control policies that minimize power consumption without sacrificing task performance. Furthermore, bio-inspired strategies—including neuromorphic computing for low-power processing and adaptive actuation mechanisms that mimic biological efficiency—present a highly promising avenue for achieving breakthrough advances in energy-efficient locomotion.

10.6. Integration with Neuroscience and Bio-Inspired Learning

Understanding biological mechanisms of learning and motor control can inspire more efficient DRL algorithms for soft robotics. Investigating biologically plausible learning frameworks, such as spiking neural networks and neuromorphic computing, may lead to breakthroughs in adaptive and intelligent soft robots [70].

10.7. Ethical and Societal Considerations

As soft mobile robots become more autonomous, ethical and societal considerations must be addressed. Future

research should explore responsible AI frameworks, transparency in decision-making, and the impact of DRL-driven automation on human labor and safety [71]. Developing ethical guidelines for DRL-based robotic systems will be essential for their widespread adoption.

11. Conclusions

This paper has provided a comprehensive survey of deep reinforcement learning techniques applied to soft mobile robotics. We have explored the integration of DRL into soft robotics, state representation, action spaces, learning strategies, simulation environments, and real-world applications. While DRL has demonstrated significant potential in enhancing the adaptability and performance of soft robots, several challenges remain, including sample efficiency, safety, sim-to-real transfer, multi-agent learning, and ethical concerns.

Future research should focus on improving DRL frameworks to achieve greater generalization, efficiency, and robustness in real-world scenarios. By addressing these challenges, DRL-driven soft robots can play a transformative role in various fields, from healthcare and industrial automation to environmental monitoring and space exploration.

Author Contributions

Conceptualization: M.H.G. and H.S.; Methodology: M.H.G.; Software: H.S.; Validation: M.H.G., H.S. and K.T.; Formal Analysis: H.S.; Investigation: H.S.; Resources: H.S.; Data Curation: M.H.G.; Writing—Original Draft Preparation: K.T.

Ethics Statement

Not applicable. This study did not involve humans or animals.

Informed Consent Statement

Not applicable. This study did not involve humans.

Data Availability Statement

Not applicable. No datasets were generated or analyzed during the current study.

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Rus D, Tolley MT. Design, fabrication and control of soft robots. *Nature* **2015**, *521*, 467–475. doi:10.1038/nature14543.
2. Kim D, Kim SH, Kim T, Kang BB, Lee M, Park W, et al. Review of machine learning methods in soft robotics. *PLoS ONE* **2021**, *16*, e0246102. doi:10.1371/journal.pone.0246102.
3. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274. doi:10.1177/0278364913495721.
4. Bhagat S, Banerjee H, Ho Tse ZT, Ren H. Deep reinforcement learning for soft, flexible robots: Brief review with impending challenges. *Robotics* **2019**, *8*, 4. doi:10.3390/robotics8010004.
5. Ijspeert AJ, Crespi A, Ryczko D, Cabelguen JM. From swimming to walking with a salamander robot driven by a spinal cord model. *Science* **2007**, *315*, 1416–1420. doi:10.1126/science.1138353.
6. Ijspeert AJ. Central pattern generators for locomotion control in animals and robots: A review. *Neural Netw.* **2008**, *21*, 642–653. doi:10.1016/j.neunet.2008.03.014.
7. Ijspeert AJ, Nakanishi J, Hoffmann H, Pastor P, Schaal S. Dynamical movement primitives: Learning attractor models for motor behaviors. *Neural Comput.* **2013**, *25*, 328–373. doi:10.1162/NECO_a_00393.

8. Wang H, Chen J, Lau HY, Ren H. Motion planning based on learning from demonstration for multiple-segment flexible soft robots actuated by electroactive polymers. *IEEE Robot. Autom. Lett.* **2016**, *1*, 391–398. doi:10.1109/LRA.2016.2521384.

9. Bern JM, Kumagai G, Coros S. Fabrication, Modeling, and Control of Plush Robots. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017. doi:10.1109/IROS.2017.8206223.

10. Zhang H, Cao R, Zilberstein S, Wu F, Chen X. Toward effective soft robot control via reinforcement learning. In *International Conference on Intelligent Robotics and Applications*; Springer International Publishing: Cham, Switzerland, 2017; pp. 173–184.

11. Nakajima K, Hauser H, Li T, Pfeifer R. Exploiting the dynamics of soft materials for machine learning. *Soft Robot.* **2018**, *3*, 339–347. doi:10.1089/soro.2017.0075.

12. Yang T, Xiao Y, Zhang Z, Liang Y, Li G, Zhang M, et al. A soft artificial muscle driven robot with reinforcement learning. *Sci. Rep.* **2018**, *8*, 14518. doi:10.1038/s41598-018-32757-9.

13. Katzschmann RK, DelPreto J, MacCurdy R, Rus D. Exploration of underwater life with an acoustically controlled soft robotic fish. *Sci. Robot.* **2018**, *3*, eaar3449. doi:10.1126/scirobotics.aar3449.

14. Thuruthel TG, Falotico E, Renda F, Laschi C. Model-based reinforcement learning for closed-loop dynamic control of soft robotic manipulators. *IEEE Trans. Robot.* **2018**, *35*, 124–134. doi:10.1109/TRO.2018.2878318.

15. Homberg B, Sanders W, Morrow J, Correll N. Robust proprioceptive grasping with a soft robot hand. *Auton. Robot.* **2018**, *43*, 681–696. doi:10.1007/s10514-018-9754-1.

16. Haarnoja T, Ha S, Zhou A, Tan J, Tucker G, Levine S. Learning to walk via deep reinforcement learning. *arXiv* **2018**, arXiv:1812.11103. doi:10.48550/arXiv.1812.11103.

17. Ishige M, Umedachi T, Taniguchi T, Kawahara Y. Exploring behaviors of caterpillar-like soft robots with a central pattern generator-based controller and reinforcement learning. *Soft Robot.* **2019**, *6*, 579–594. doi:10.1089/soro.2018.0126.

18. Satheeshbabu S, Uppalapati NK, Chowdhary G, Krishnan G. Open loop position control of soft continuum arm using deep reinforcement learning. *IEEE Robot. Autom. Lett.* **2019**, 5133–5139. doi:10.1109/ICRA.2019.8793653.

19. Park T, Cha Y. Soft mobile robot inspired by animal-like running motion. *Sci. Rep.* **2019**, *9*, 14700. doi:10.1038/s41598-019-51308-4.

20. Truby RL, Katzschmann RK, Lewis JA, Rus D. Soft Robotic Fingers with Embedded Ionogel Sensors and Discrete Actuation Modes for Somatosensi ive Manipulation. In Proceedings of the 2019 2nd IEEE International Conference on Soft Robotics (RoboSoft), Seoul, Republic of Korea, 14–18 April 2019; IEEE: New York, NY, USA, 2019; pp. 322–329. doi:10.1109/ROBOSOFT.2019.8722722.

21. Berndt A, Pfaff P, Ryll A, Allaire PE. Soft Robot Control with a Learned Differentiable Model. In Proceedings of the 2020 3rd IEEE International Conference on Soft Robotics (RoboSoft), New Haven, CT, USA, 15 May–15 July 2020; IEEE: New York, NY, USA, 2020; pp. 417–423. doi:10.1109/RoboSoft48309.2020.9116011.

22. Li G, Shintake S, Shintake M. Deep Reinforcement Learning Framework for Underwater Locomotion of Soft Robot. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2021), Xi'an, China, 30 May–5 June 2021; pp. 12033–12039. doi:10.1109/ICRA48506.2021.9561145.

23. Raeisinezhad M, Pagliocca N, Koohbor B, Trkov M. Design optimization of a pneumatic soft robotic actuator using model-based optimization and deep reinforcement learning. *Front. Robot. AI* **2021**, *8*, 639102. doi:10.3389/frobt.2021.639102.

24. Yin X, Müller R. Integration of deep learning and soft robotics for a biomimetic approach to nonlinear sensing. *Nat. Mach. Intell.* **2021**, *3*, 507–512. doi:10.1038/s42256-021-00330-1.

25. Thanabalan HP. Learning Soft Robot and Soft Actuator Dynamics Using Deep Neural Network. Master's Thesis, Queen Mary University of London, London, UK, 2020.

26. Lin YH, Siddall R, Schwab F, Fukushima T, Banerjee H, Baek Y, et al. Modeling and Control of a Soft Robotic Fish with Integrated Soft Sensing. *Adv. Intell. Syst.* **2023**, *5*, 2000244. doi:10.1002/aisy.202000244.

27. Wang Z, Li X, Cheng L, Shen Y. Review on Reinforcement Learning Controller in Soft Manipulator. In Proceedings of the 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 27–31 December 2021; IEEE: New York, NY, USA, 2021; pp. 558–563. doi:10.1109/ROBIO54168.2021.9739342.

28. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1999; Volume 17, pp. 229–235.

29. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. doi:10.1038/nature14236.

30. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347. doi:10.48550/arXiv.1707.06347.

31. Fujimoto S, Hoof H, Meger D. Addressing Function Approximation Error in Actor-Critic Methods. In *International Conference on Machine Learning*; PMLR: Norfolk, MA, USA, 2018; pp. 1587–1596.

32. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971. doi:10.48550/arXiv.1509.02971.

33. Schulman J, Levine S, Moritz P, Jordan MI, Abbeel P. Trust Region Policy Optimization (Trpo). In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015.

34. Rajeswaran A, Kumar V, Gupta A, Vezzani G, Schulman J, Todorov E, et al. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. *arXiv* **2017**, arXiv:1709.10087. doi:10.48550/arXiv.1709.10087.

35. Duriez C. Control of Elastic Soft Robots Based on Real-Time Finite Element Method. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; IEEE: New York, NY, USA, 2013; pp. 3982–3987. doi:10.1109/ICRA.2013.6631138.

36. Todorov E, Erez T, Tassa Y. Mujoco: A Physics Engine for Model-Based Control. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Portugal, 7–12 October 2012; IEEE: New York, NY, USA, 2012; pp. 5026–5033.

37. Finn C, Abbeel P, Levine S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; PMLR: Norfolk, MA, USA, 2017; Volume 70, pp. 1126–1135.

38. Peng XB, Abbeel P, Levine S, van de Panne M. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–14. doi:10.1145/3197517.3201311.

39. Tzeng E, Devin C, Hoffman J, Finn C, Abbeel P, Levine S, et al. Adapting deep visuomotor representations with weak pairwise constraints. In *Algorithmic Foundations of Robotics XII: Proceedings of the Twelfth Workshop on the Algorithmic Foundations of Robotics*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 688–703. doi:10.1007/978-3-030-43089-4-44.

40. Levine S, Finn C, Darrell T, Abbeel P. End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.* **2016**, *17*, 1–40.

41. Kumar S, Rani K, Banga V. Robotic arm movement optimization using soft computing. *IAES Int. J. Robot. Autom. (IJRA)* **2017**, *6*, 1–14. doi:10.11591/ijra.v6i1.pp1-14.

42. Qin L, Liang X, Huang H, Chui CK, Yeow RCH, Zhu J. A versatile soft crawling robot with rapid locomotion. *Soft Robot.* **2019**, *6*, 455–467. doi:10.1089/soro.2018.0124.

43. Botteghi N, Obbink R, Geijs D, Poel M, Sirmacek B, Brune C, et al. Low Dimensional State Representation Learning with Reward-Shaped Priors. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE: New York, NY, USA, 2021; pp. 3736–3743.

44. Miriyev A, Stack K, Lipson H. Soft material for soft actuators. *Nat. Commun.* **2017**, *8*, 596. doi:10.1038/s41467-017-00685-3.

45. Merckling A, Perrin-Gilbert N, Coninx A, Doncieux S. Exploratory state representation learning. *Front. Robot. AI* **2022**, *9*, 762051. doi:10.3389/frobt.2022.762051.

46. Zhu J, Navarro B, Passama R, Fraisse P, Crosnier A, Cherubini A. Robotic manipulation planning for shaping deformable linear objects with environmental contacts. *IEEE Robot. Autom. Lett.* **2019**, *5*, 16–23. doi:10.1109/LRA.2019.2944304.

47. Majidi C. Soft robotics: A perspective—current trends and prospects for the future. *Soft Robot.* **2014**, *1*, 5–11. doi:10.1089/soro.2013.0001.

48. Laschi C, Cianchetti M, Mazzolai B, Margheri L, Follador M, Dario P. Soft robot arm inspired by the octopus. *Adv. Robot.* **2012**, *26*, 709–727 doi:10.1163/156855312X626343.

49. Chen T, Ben A, Jiaheng H, Rohan C, Roberto M-M, Peter S. Deep reinforcement learning for robotics: A survey of real-world successes. *Annu. Rev. Control. Robot. Auton. Syst.* **2025**, *26*, 153–188. doi:10.1146/annurev-control-030323-022510.

50. Janner M, Fu J, Zhang M, Levine S. When to Trust your Model: Model-Based Policy Optimization. Part of Advances in Neural Information Processing Systems. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Volume 32.

51. He L, Qi X, Chen D, Li Z, Tan X. DiSA-IQL: Offline Reinforcement Learning for Robust Soft Robot Control under Distribution Shifts. *arXiv* **2025**, arXiv:2510.00358. doi:10.48550/arXiv.2510.00358.

52. Cianchetti M, Laschi C, Menciassi A, Dario P. Biomedical applications of soft robotics. *Nat. Rev. Mater.* **2018**, *3*, 143–153. doi:10.1038/s41578-018-0022-y.

53. Trivedi D, Rahn CD, Kier WM, Walker ID. Soft robotics: Biological inspiration, state of the art, and future research. *Appl. Bionics Biomech.* **2008**, *5*, 99–117. doi:10.1080/11762320802557865.

54. Lee D. Reinforcement Learning Techniques for Autonomous Robots in Virtual Environments with LLM-Based Multimodal Data Integration and Virtual Embodiment. *Clin. Res. Clin. Trials* **2025**, *12*. doi:10.31579/2693-4779/263.

55. Naughton N, Sun J, Tekinalp A, Parthasarathy T, Chowdhary G, Gazzola M. Elastica: A compliant mechanics environment for soft robotic control. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3389–3396. doi:10.1109/LRA.2021.3063698.

56. Peng XB, Andrychowicz M, Zaremba W, Abbeel P. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; IEEE: New York, NY, USA, 2018; pp. 3803–3810. doi:10.1109/ICRA.2018.8460528.

57. James S, Wohlhart P, Kalakrishnan M, Kalashnikov D, Irpan A, Ibarz J, et al. Sim-to-Real via Sim-to-Sim: Data-Efficient Robotic Grasping via Randomized-to-Canonical Adaptation Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12627–12637.

58. Hwangbo J, Lee J, Dosovitskiy A, Bellicoso D, Tsounis V, Koltun V, et al. Learning agile and dynamic motor skills for legged robots. *Sci. Robot.* **2019**, *4*, eaau5872. doi:10.1126/scirobotics.aau5872.

59. Shintake J, Cacucciolo V, Shea H, Floreano D. Soft biomimetic fish robot made of dielectric elastomer actuators. *Soft Robot.* **2018**, *5*, 466–474. doi:10.1089/soro.2017.0062.

60. Della Santina C, Catalano MG, Bicchi A. Soft robots. In *Encyclopedia of Robotics*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–14.

61. Kim KJ, Tadokoro S. *Electroactive Polymers for Robotic Applications: Artificial Muscles and Sensors*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 23, p. 291.

62. Shepherd RF, Ilievski F, Choi W, Morin SA, Stokes AA, Mazzeo AD, et al. Multigait soft robot. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20400–20403. doi:10.1073/pnas.1116564108.

63. Zhao H, O'Brien K, Li S, Shepherd RF. Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides *Sci. Robot.* **2016**, *1*, eaai7529. doi:10.1126/scirobotics.aai7529.

64. Marchese AD. Design, Fabrication, and Control of Soft Robots with Fluidic Elastomer Actuators. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2015.

65. Li L, Donato E, Lomonaco V, Falotico E. Continual Policy Distillation of Reinforcement Learning-Based Controllers for Soft Robotic In-Hand Manipulation. In Proceedings of the 2024 IEEE 7th International Conference on Soft Robotics (RoboSoft), San Diego, CA, USA, 14–17 April 2024; IEEE: New York, NY, USA, 2024; Volume 37, pp. 1026–1033. doi:10.1109/RoboSoft60065.2024.10522027.

66. Nagabandi A, Finn C, Levine S. Deep online learning via meta-learning: Continual adaptation for model-based rl. *arXiv* **2018**, arXiv:1812.07671. doi:10.48550/arXiv.1812.07671.

67. García J, Fernández F. A comprehensive survey on safe reinforcement learning. *J. Mach. Learn. Res.* **2015**, *16*, 1437–1480.

68. Ding Z, Tsai Y-Y, Lee WW, Huang B. Sim-to-Real Transfer for Robotic Manipulation with Tactile Sensory. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: New York, NY, USA, 2021; pp. 6778–6785. doi:10.48550/arXiv.2103.00410.

69. Lowe R, Wu YI, Tamar A, Harb J, Pieter A, Mordatch I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.

70. Krichmar JL, Chou T-S. A tactile robot for developmental disorder therapy. In *Proceedings of the Technology, Mind, and Society*; TechMindSociety: Online, 2018; pp. 1–6. doi:10.1145/3183654.3183657.

71. Bryson JJ. Patiency is not a virtue: The design of intelligent systems and systems of ethics. *Ethics Inf. Technol.* **2018**, *20*, 15–26. doi:10.1007/s10676-018-9448-6.