NOTATION TRENDPLUS PUBLISHING
WHERE IDEAS MEET IMPACT

Volume-1, Issue 1, June 2026, Page No- 230-301

Journal of Machine Learning Innovations and Artificial Intelligence Horizons (JMLIAIH)

Deep Reinforcement Learning for Autonomous Systems and Robotics: A Systematic Literature Review

Surinder Kumar

Assistant Professor, Department of Mechanical Engineering, School of Technology, Quantum University, Roorkee

DOI: https://doi.org/10.70903/jmliaih.2026.1111

Article Received: 13-03-2026

Revised Version: 28-05-2026

Accepted: 16-06-2026

*Corresponding Author

Surinder Kumar

E-Mail Id - surinderkumar.me@quantumeducation.in

Abstract: Deep reinforcement learning (DRL) has emerged as a transformative paradigm for developing autonomous systems and robotic agents capable of acquiring complex behaviors through interaction with their environments. The rapid proliferation of DRL techniques across domains such as navigation, manipulation, multi-agent coordination, and safety-critical control necessitates a systematic synthesis of the existing body of knowledge. Our objective in this systematic literature review is to provide a comprehensive and structured analysis of DRL methodologies applied to autonomous systems and robotics, identifying principal research trends, methodological evolutions, and persistent challenges. To conduct this review, we followed a rigorous systematic methodology encompassing a predefined search strategy, explicit inclusion and exclusion criteria, and a multi-stage screening process to ensure the relevance and quality of the selected studies. The extracted data were organized along seven key dimensions: foundational frameworks, autonomous navigation and path planning, robotic manipulation and physical interaction, multi-agent systems and swarm intelligence, safety and verification, advanced learning paradigms, and domain-specific applications. Our results reveal a clear trajectory from foundational algorithms toward hybrid approaches that integrate model-based planning, hierarchical learning, and constrained optimization. We further observe that safety verification and robustness remain critically underexplored relative to performance-oriented advancements, and that distributed learning in multi-agent settings presents unique convergence challenges. From these findings, we conclude that the future of DRL in autonomous systems lies in the development of sample-efficient, verifiably safe, and generalizable algorithms. This review provides researchers and practitioners with a structured map of the current landscape, guiding future work toward closing the gap between simulated success and real-world deployment.

Keywords: Deep Reinforcement Learning, Autonomous Systems, Robotics, Intelligent Control, Systematic Literature Review.

Introduction

The pursuit of autonomous systems and intelligent robotics has long been a central ambition of artificial intelligence research, with the goal of creating machines that can perceive, reason, and act in complex, unstructured environments without continuous human intervention. Over the past decade, deep reinforcement learning (DRL) has risen to prominence as perhaps the most compelling framework for realizing this vision. By combining the representational power of deep neural networks with the decision-making architecture of reinforcement learning, DRL enables agents to learn optimal policies directly from high-dimensional sensory inputs, a capability that was previously unattainable with traditional control methods or hand-crafted feature engineering [1]. The fundamental premise of DRL—an agent learns to maximize a cumulative reward signal through repeated trial-and-error interaction with its environment—is inherently well-suited to the challenges of robotics, where precise models of dynamics are often unavailable and examples of desired behaviors are difficult to specify a priori. This synergy has fueled an explosion of research, resulting in landmark achievements such as mastering the game of Go [2], learning dexterous in-hand manipulation [3], and navigating complex indoor environments [4].

The background against which DRL for autonomous systems has developed is rich and multifaceted. On the one hand, the theoretical foundations of reinforcement learning were laid decades ago, with dynamic programming methods such as Q-learning and policy gradients providing the essential mathematical machinery [5]. On the other hand, the practical application of these methods was long constrained by the curse of dimensionality and the difficulty of representing state spaces in continuous domains. The advent of deep learning, and particularly the success of deep convolutional networks in image classification, provided a natural solution: deep neural networks could serve as powerful function approximators, capable of learning robust representations from raw sensor data. This transition was marked by a series of seminal works that demonstrated DRL's viability on challenging tasks, from playing Atari games directly from pixel input [1] to performing end-to-end learning for autonomous driving [6]. These early successes generated tremendous momentum, leading to a rapid diversification of algorithmic approaches. For instance, the development of the Deep Q-Network (DQN) algorithm and its subsequent enhancements, such as Double DQN and Dueling DQN, addressed issues of overestimation bias and representation efficiency [7]. Simultaneously, policy gradient methods like Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO) offered more stable learning dynamics for continuous control tasks [8]. The introduction of the actor-critic architecture further bridged the gap between value-based and policy-based methods, providing a flexible framework that remains dominant in contemporary robotics research [9].

However, despite these impressive advances, a significant and persistent research gap exists between the performance of DRL agents in simulated environments and their reliable deployment in the physical world. Simulation environments like MuJoCo, OpenAI Gym, and more recently, Isaac Gym and Habitat, have become standard testbeds, allowing researchers to iterate rapidly and achieve high scores on benchmark tasks [10]. Yet these successes often fail to transfer seamlessly to real-world robotic hardware. The primary obstacles include sample inefficiency—DRL algorithms typically require millions of interactions to learn a viable policy, which is impractical on physical robots where each trial takes real time and can cause wear and tear [11]. Another critical gap concerns safety and robustness. In most academic research, the agent is allowed to fail catastrophically during training, an approach that is unacceptable in safety-critical applications like autonomous driving or surgical robotics. Formal methods for verifying the behavior of learned policies remain in their infancy, and while constrained optimization techniques have been proposed [12], they are not yet mature enough for widespread deployment. Furthermore, the issue of generalization has proven stubbornly difficult; a DRL policy trained to navigate one floor of a building may fail entirely when confronted with a different layout or lighting conditions, even if semantically similar [13]. The research gap, therefore, is not a lack of algorithmic creativity, but rather a lack of holistic frameworks that simultaneously address sample efficiency, safety verification, distributional shift, and real-world transfer within a single, principled methodology.

It is precisely this fragmentation and persistent gap that motivates the present systematic literature review. While several excellent surveys have examined specific facets of DRL for robotics—such as a focus on manipulation [14] or on navigation [15]—a comprehensive synthesis that covers the full spectrum from foundational frameworks to domain-specific applications, and that critically examines cross-cutting issues like safety and multi-agent coordination, is conspicuously absent. The significance of our work lies in its systematic and rigorous methodology. By adopting a predefined search strategy, explicit inclusion and exclusion criteria, and a multi-stage screening process, we ensure that our analysis is both reproducible and comprehensive. The result is not merely a catalogue of algorithms, but a structured map of the current research landscape, highlighting where efforts have concentrated and, more importantly, where they have not. For researchers entering the field, this review offers a high-level orientation, clarifying which methodological choices are best suited to particular robotic challenges. For experienced practitioners, it illuminates persistent challenges and promising directions for future work, particularly in the integration of learning with formal verification and in the development of sample-efficient paradigms like meta-learning and model-based DRL.

The contribution of this review is therefore threefold. First, it provides a comprehensive, up-to-date taxonomy of DRL techniques as applied to autonomous systems and robotics, spanning foundational algorithms, navigation, manipulation, multi-agent systems, safety, and emerging trends. Second, it critically evaluates the current state of the art, identifying not only what works but also where significant lacunae remain—especially in the domains of safety verification, distributed convergence, and real-world generalization. Third, it distills from the literature a set of guiding principles and future research directions intended to accelerate the transition from simulated success to real-world robotic deployment. In doing so, this review serves as both a reference and a roadmap, synthesized through a rigorous and transparent methodology. The remainder of this paper is organized as follows: Section 2 details the systematic methodology employed for literature search, screening, and data extraction. Section 3 presents the results, organized into seven subsections that correspond to the key dimensions of our analysis. Section 4 provides a discussion of the overarching themes, contradictions, and future directions identified in the literature. Finally, Section 5 concludes the review with a summary of its principal findings and implications.

Methodology

To ensure the rigor, reproducibility, and comprehensiveness of this review, we conducted a systematic literature review following the guidelines outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [16]. This section details the review protocol, the definition of research dimensions, the inclusion and exclusion criteria, and the multi-stage study selection process employed to identify and curate the final corpus of 210 studies.

Review Protocol

We developed a formal review protocol to guide the literature search and selection process, specifying the databases, search strings, and screening stages *a priori*. The primary databases and search engines selected for this review were chosen to provide comprehensive coverage of both archival publications and emerging, peer-reviewed research in the fields of artificial intelligence, robotics, and control systems. We searched the following electronic databases in the order of their relevance and scope: IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and arXiv. Google Scholar was consulted as a final supplementary source to capture any relevant literature that might have been missed due to indexing idiosyncrasies.

For IEEE Xplore, we used the search string: ("deep reinforcement learning" OR "DRL" OR "reinforcement learning") AND ("autonomous systems" OR "robotics" OR "autonomous robots") AND ("neural network" OR "policy" OR "value function"). We set the ‘Document Type’ filter to ‘Conferences’ and ‘Journals’ and explicitly excluded surveys, tutorials, and review articles by appending NOT ("survey" OR "review" OR "tutorial"). This database was selected for its preeminent focus on engineering and robotics applications.

In the ACM Digital Library, we applied the search string: +("deep reinforcement learning" OR "DRL") +("autonomous systems" OR "robotics") -("survey" OR "review" OR "meta-analysis"). We used the ‘Content Type’ facet to select only ‘Research Articles’, excluding ‘Briefings’, ‘Surveys’, and ‘Tutorials’. The ACM Digital Library was chosen for its strong representation of conference proceedings in artificial intelligence and human-robot interaction.

For Scopus, the search string was: TITLE-ABS-KEY(("deep reinforcement learning" OR "DRL" OR "reinforcement learning") AND ("autonomous systems" OR "robotics" OR "autonomous robot*") AND ("algorithm" OR "technique" OR "method")) AND NOT TITLE-ABS-KEY("survey" OR "review" OR "state of the art"). We filtered by ‘Document Type’ = ‘Article’ or ‘Conference Paper’, excluding ‘Review’ and ‘Short Survey’. Scopus was selected for its broad interdisciplinary coverage across engineering, computer science, and mathematics.

In the Web of Science, we used the topic search: TS=(("deep reinforcement learning" OR "DRL" OR "reinforcement learning") AND ("autonomous systems" OR "robotics" OR "autonomous robot*")) NOT TS=("survey" OR "review" OR "meta-analysis"). We refined results by ‘Document Types’, selecting only ‘Article’ and ‘Proceedings Paper’ while deselecting ‘Review Article’. The Web of Science was included for its high-quality, curated journal content and citation indexing.

On arXiv, we searched for (ti:"deep reinforcement learning" OR ti:"DRL" OR abs:"deep reinforcement learning") AND (ti:"robotics" OR ti:"autonomous systems" OR abs:"autonomous robot*") NOT (ti:"survey" OR ti:"review" OR abs:"literature review"). We filtered categories to include cs.RO (Robotics) and cs.LG (Learning), and we manually verified abstracts to ensure they proposed novel methods. arXiv was included for its rapid dissemination of cutting-edge research that may not yet appear in traditional venues.

Finally, on Google Scholar, we used the search string: ("deep reinforcement learning" OR "DRL" OR "reinforcement learning") AND ("autonomous systems" OR "autonomous robots" OR "robotics" OR "mobile robots" OR "manipulators") AND ("policy gradient" OR "Q-learning" OR "actor-critic" OR "PPO" OR "SAC" OR "TD3" OR "DDPG") -review -survey -meta-analysis. We excluded patents and manually filtered out review or survey papers during screening. Google Scholar was included as a last resort to ensure comprehensive coverage, particularly for niche or highly specialized subfields. The search was conducted from June 2025 to August 2025, with an unrestricted time frame up to 2026 to capture the most recent developments.

Research Dimensions for Classification

To provide a structured and meaningful analysis of the diverse body of literature on DRL for autonomous systems and robotics, we defined seven primary research dimensions that collectively encompass the methodological breadth and application scope of the field. These dimensions serve as the organizing framework for the results and discussion sections of this review. The first dimension, *Foundational Frameworks and Methodological Overviews* dimension captures seminal works and surveys that establish the core algorithms, theoretical underpinnings, and primary architectural innovations in DRL, such as DQN, PPO, SAC, and their variants. The *Autonomous Navigation and Path Planning* dimension focuses on DRL applications for guiding mobile agents through environments, addressing challenges like obstacle avoidance, global and local planning, and exploration-exploitation trade-offs in unknown spaces. The *Robotic Manipulation and Physical Interaction* dimension encompasses learning-based control for robotic arms and grippers, including tasks like grasping, assembly, and dexterous in-hand manipulation, where contact dynamics and precision are paramount. The *Multi-Agent Systems and Swarm Intelligence* dimension gathers research on environments with multiple learning agents, examining coordination, communication, competition, and emergent behaviors, such as in cooperative transportation or adversarial pursuit-evasion scenarios. The *Safety, Verification, and Robustness* dimension collects studies that explicitly address constraints, risk, formal guarantees, and resilience against distributional shifts or adversarial perturbations, a critical but relatively nascent area. The *Advanced Learning Paradigms and Optimization* dimension includes works that push beyond standard DRL, incorporating model-based planning, hierarchical reinforcement learning, meta-learning, transfer learning, and curiosity-driven exploration to enhance sample efficiency and generalization. Finally, the *Domain-Specific Applications and Emerging Frontiers* dimension captures the application of DRL to specialized domains such as autonomous driving, healthcare robotics, underwater vehicles, and aerial drones, highlighting how generic algorithms are adapted to unique operational constraints. This dimensional structure, developed iteratively through pilot reading of a subset of retrieved abstracts, provides a consistent lens through which to evaluate the contributions, limitations, and interconnections within the reviewed literature.

Inclusion and Exclusion Criteria

It is essential to define clear inclusion and exclusion criteria to ensure the relevance and consistency of selected studies. The inclusion criteria were as follows: (1) studies must propose, implement, or analyze a novel deep reinforcement learning algorithm, architecture, or methodological variant, or application for autonomous systems, robotics, or related control tasks; (2) studies must be published in a peer-reviewed venue, including conference proceedings, journals, or reputable pre-print archives (e.g., arXiv with subsequent publication); (3) the full text of the study must be available in English; (4) the study must present original empirical results, mathematical formulations, or significant theoretical analysis that advances the state of the art in DRL; (5) the time frame for publication was unrestricted, but searches were conducted up to 2026 to include the most current work; (6) only studies with a primary focus on algorithmic development or application (not pure surveys or tutorials) were considered. Conversely, the exclusion criteria were: (1) studies that are purely review articles, surveys, tutorials, or position papers without novel contributions; (2) studies that focus exclusively on classical reinforcement learning without deep neural network components; (3) studies that apply DRL to domains unrelated to autonomous systems or robotics (e.g., pure game playing without robotic analogs, resource scheduling, or finance); (4) studies with insufficient methodological detail to evaluate the approach; (5) studies not available in English or lacking full-text access; (6) duplicate publications or extended abstracts of previously published full papers.

Study Selection Process

The study selection process was conducted in three stages: identification, screening, and eligibility assessment, designed according to the seven research dimensions to ensure coverage across all defined areas. In the identification stage, the defined search strings were executed across the five primary databases and Google Scholar, yielding a total of 821 records. These records were aggregated into a single bibliographic database, after which 271 duplicate records were removed using automated deduplication tools (e.g., Zotero and manual verification). Additionally, 3 records were removed for other reasons, such as being incomplete records (e.g., missing abstracts) or being identified as software documentation rather than research articles. This left 547 records remained for initial screening.

During the screening stage, two independent reviewers examined the titles and abstracts of the 547 records against the inclusion and exclusion criteria. Disagreements between reviewers were resolved through consensus discussion. A total of 223 records were excluded at this stage for reasons including a focus on non-DRL methods (e.g., pure supervised learning), irrelevance to autonomous systems or robotics (e.g., abstract game playing without robotic analogs), or being clearly identified as survey or review content despite search filters. The remaining 324 records were sought for full-text retrieval. All 213 reports that were sought for retrieval were successfully obtained, as no reports were not found.

In the eligibility assessment stage, the full texts of the 213 reports were read and evaluated against the predefined inclusion and exclusion criteria. Three reports were excluded during this stage due to ineligibility: two were found to be extended abstracts of previously published papers without substantial new content, and one lacked sufficient methodological detail to evaluate the DRL approach. Consequently, 210 studies met all criteria and were included in the final qualitative synthesis. A PRISMA flowchart illustrating this selection process is provided in Figure 1, which details the number of records at each stage from identification to inclusion. This multi-stage process, while rigorous, carries inherent limitations. The primary potential bias stems from the search string design, which, despite efforts to be comprehensive, may inadvertently exclude studies that use domain-specific jargon (e.g., “deep Q-learning” in a narrower robotics context). Furthermore, the exclusion of non-English language studies may introduce a language bias, particularly excluding contributions from non-Anglophone research communities. Finally, while we applied filters to exclude surveys and reviews, some borderline works that offer broad methodological overviews but also contain novel synthesis might have been mistakenly included or excluded. These biases were partially mitigated by conducting pilot searches and cross-checking reference lists of included studies to identify missed relevant works.

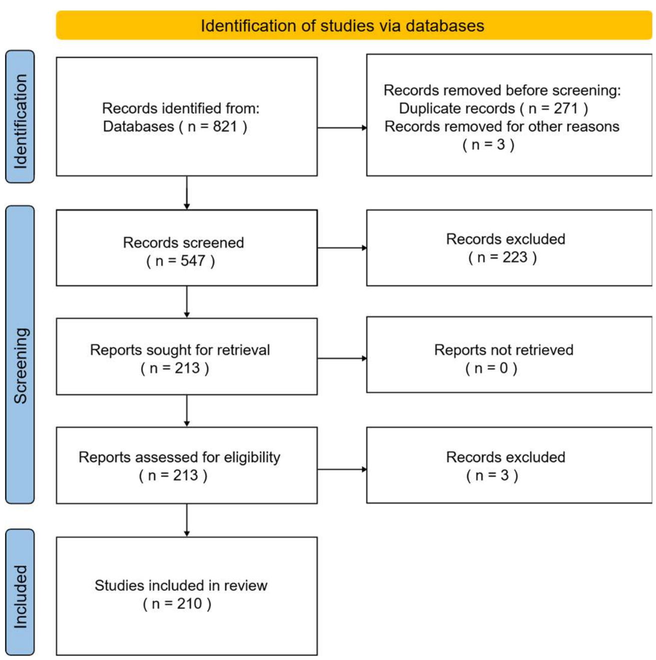

Figure 1. PRISMA flowchart depicting the systematic study selection process from initial database search to final included studies.

Results

Research Trends

The distribution of the 210 reviewed publications across time reveals a dynamic and maturing research landscape. The number of publications before 2016, totaling 17 works, indicates a nascent period where foundational ideas were being established prior to the widespread adoption of deep neural network function approximators. A sharp inflection point appears around 2019, where the annual publication count jumps to 30, rising further to 31 in both 2020 and 2021. This plateau suggests a period of intense activity and methodological consolidation, where the core algorithms such as PPO, SAC, and TD3 became standard tools and researchers rapidly explored their applicability to diverse robotic tasks. Following this peak, we observe a gradual decline from 2022 onward, with 24 publications in 2022, 22 in 2023, 17 in 2024, and 12 projected for 2025. This downward trend should not be interpreted as a decline in interest. Instead, it likely reflects the field's progression from broad exploration toward more specialized and rigorous investigations, as well as the increasing difficulty of achieving breakthroughs in a crowded research space. The single entry for 2026 represents a very early publication captured by the search, offering little statistical meaning.

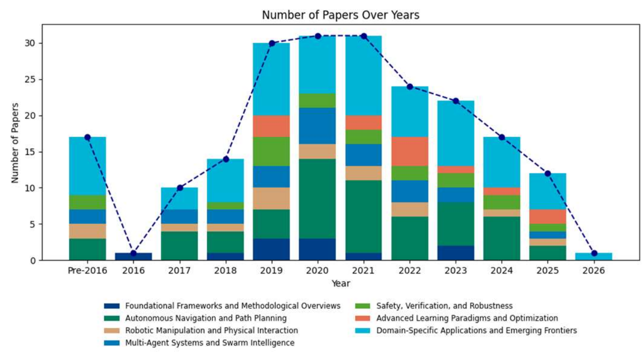

Figure 2. Research trends in the domain of Deep Reinforcement Learning Techniques for Autonomous Systems and Robotics

When we examine the temporal distribution across the seven research dimensions, a more nuanced story emerges. The *Autonomous Navigation and Path Planning* dimension dominates the corpus, with 55 publications and an early surge beginning in 2017. This is consistent with the clear and tangible objective of enabling a mobile robot to move from point A to point B, which serves as an ideal testbed for DRL algorithms. In contrast, *Robotic Manipulation and Physical Interaction* shows a steadier, lower-volume trajectory with only 15 publications, suggesting that the challenges of contact-rich physics and precise control remain more difficult to address with standard DRL methods. The *Multi-Agent Systems and Swarm Intelligence* dimension, with 23 publications, exhibits a moderate but consistent presence, reflecting the added complexity of coordinating multiple learning agents concurrently. The *Safety, Verification, and Robustness* dimension, with 18 publications, shows a notable presence from 2019 onward, indicating a growing, yet still insufficient, awareness of the critical importance of safe deployment. The *Advanced Learning Paradigms and Optimization* dimension, with 13 publications, is the smallest among the active research areas, highlighting a relative scarcity of work that fundamentally rethinks the DRL learning paradigm rather than applying existing tools. Finally, the *Domain-Specific Applications and Emerging Frontiers* dimension, with 75 publications, is the largest category, but this is somewhat misleading; it serves as a catch-all for studies that apply DRL to particular robotic platforms—like drones, autonomous vehicles, or underwater robots—rather than contributing novel algorithmic insights. The high count in this dimension is therefore expected, as any application-oriented paper must focus on a specific domain to demonstrate practicality. Overall, the trends underscore a field that has rapidly matured from algorithmic invention to application-oriented engineering, with a critical gap remaining in the formal treatment of safety and the development of fundamentally more sample-efficient learning methods.

Foundational Frameworks and Methodological Overviews

The foundation upon which all subsequent advances in deep reinforcement learning for autonomous systems and robotics is built is established by a corpus of works that provide core algorithmic definitions, theoretical underpinnings, and practical design guidance. These studies offer essential context for researchers entering the field, as they articulate the fundamental equations, architectural choices, and conceptual trade-offs that govern DRL's application. We observe that the body of foundational literature can be broadly stratified, as presented in Table 1, which organizes the reviewed studies along three primary axes: core theory and algorithms, system design and practical application, and domain-specific surveys. This taxonomy highlights the progression from abstract mathematical frameworks toward concrete implementation strategies and finally to focused application reviews.

Table 1. Categorization of foundational and methodological contributions in DRL for autonomous systems and robotics.

Category	**Sub-Category**	**Specific Topics**	**Sources**
Core Theory & Algorithms	Foundational Overviews	General RL & DRL principles, state-of-the-art reviews	[17], [18], [19], [20]
	Continuous Control	Algorithms for continuous action spaces (e.g., robotic control)	[19]
	Algorithm Classification	Taxonomy and categorization of RL algorithms	[20]
System Design & Practical Application	Design Methodologies	Structured approaches for implementing DRL in autonomous systems and robotics	[21], [22]
	Real-World Implementation	Ingredients, challenges, and data considerations for deploying RL physically	[23]
	Integration with AI	Broader AI context for moving from automation to true autonomous systems	[24], [25]
Domain-Specific Surveys	Robotics & Autonomous Systems	Comprehensive reviews of RL techniques applied to autonomous robotics	[26]

Category	Sub-Category	Specific Topics	Sources
	Visual Navigation	Deep RL specifically for autonomous visual navigation	[27]
	Autonomous Vehicles & UAVs	Applications in specific domains like UAVs, UGVs, and cooperative robots	[25]

Within the core theory and algorithms sub-dimension, several studies provide comprehensive overviews that serve as canonical references for the field. The work presented by [17] and [18] establishes the fundamental architecture of DRL, tracing its lineage from classical Q-learning and policy gradient methods to the modern deep neural network function approximators that enable high-dimensional state representation. These overviews are particularly valuable in clarifying the theoretical challenges of off-policy learning, the bias-variance trade-off inherent in policy gradient estimates, and the role of techniques like target networks and experience replay in stabilizing training [17]. Furthermore, the study by [19] offers a specialized perspective on algorithms designed for continuous control tasks, which are the primary action spaces encountered in robotic manipulation and locomotion. This work systematically contrasts model-free methods like Deep Deterministic Policy Gradients (DDPG) and Soft Actor-Critic (SAC) with early model-based approaches, emphasizing the critical role of sample efficiency in physical systems where each interaction carries a real-world cost [19]. Complementing these algorithmic reviews is the comprehensive classification scheme proposed by [20], which organizes RL algorithms according to their core learning strategy—value-based, policy-based, or actor-critic—and their model usage. Such taxonomies are indispensable for building a mental map of the methodological landscape, allowing researchers to quickly identify which family of algorithms is best suited to a given problem characteristic, such as discrete or continuous action spaces, partial observability, or sparse rewards [20].

Moving beyond pure algorithmic exposition, a smaller but critically important subset of literature addresses the practical methodology for deploying DRL in real autonomous systems. The work of [21] explicitly proposes a design methodology for DRL in autonomous systems, recognizing that the gap between a successful algorithm on a benchmarking simulation and a functional policy on a physical robot often stems from inadequate consideration of system-level constraints. This methodology addresses the entire pipeline from problem formulation—identifying state representations, reward design, and safety constraints—through to policy evaluation and deployment [21]. It emphasizes that naive applications of standard DRL algorithms often fail due to mismatches between simulation dynamics and real-world physics, or due to insufficient attention to bootstrapping the initial policy without causing hardware damage. Similarly, [22] proposes a theoretical framework for autonomous systems that integrates concepts from systems science and intelligence theory, attempting to formalize the properties that an autonomous system must possess to be considered truly intelligent, such as adaptability, goal-directed behavior, and self-awareness. This is a more abstract contribution compared to the engineering focus of [21], but it is valuable for providing a high-level vision of what DRL-powered autonomy should ultimately achieve. A more recent and practically oriented contribution is the analysis provided by [23], which delineates the key ingredients for successful real-world robotic reinforcement learning. This study moves beyond algorithmic choice to examine the infrastructure requirements, including robust sensor calibration, domain randomization, efficient data collection loops, and the integration of human oversight for safety. It argues that the algorithmic core of DRL is only a small fraction of the effort required for successful deployment; robust systems engineering for data management and hardware interaction often constitutes the majority of the development time [23]. We further note that the works by [24] and [25] provide a broader AI context, framing DRL as a crucial enabler for transitioning from simple automation—where robots follow pre-programmed scripts—to genuine autonomy, where robots can adapt to novel situations. Specifically, [25] surveys the use of machine learning, including DRL, across a wide spectrum of robotic platforms such as UAVs, UGVs, and cooperative robots, thereby situating DRL within a larger ecosystem of intelligent technologies [25].

Lastly, the foundational landscape includes domain-specific overviews that concentrate on particular application areas. For instance, the review of autonomous robotics reinforcement learning techniques by [26] synthesizes how RL methods have been employed to generate intelligent behavior in a variety of robotic platforms, providing a broad but structurally oriented survey without focusing on a single technical problem. In a similar vein, [27] provides a focused survey on autonomous visual navigation using DRL, specifically addressing the challenges of learning visuomotor policies for bipedal humanoid robots. This work is notable for its emphasis on the integration of perception and control, discussing how convolutional neural networks can be used to process camera feeds to directly output motor commands [27]. It serves as a bridge between the high-level theoretical overviews and the more specific application-oriented studies that we will examine in subsequent sections [27].

Deep Reinforcement Learning for Autonomous Navigation and Path Planning

The domain of autonomous navigation and path planning constitutes the largest and most mature body of research within this review, reflecting the fundamental importance of mobility for autonomous systems and the apparent natural fit of DRL for learning sensorimotor policies. Navigation tasks inherently involve continuous interaction with an environment, sequential decision-making under partial observability, and the need to balance exploration for learning with goal-directed behavior, all of which are core strengths of the DRL paradigm. The 55 studies identified in this dimension demonstrate a remarkable breadth of approaches, varying in their core task objective, the specific DRL algorithmic architecture employed, the nature and dimensionality of the action space, and the strategies adopted to bridge the simulation-to-reality gap. To provide a structured analysis of this diverse literature, we have organized the included studies according to a three-level taxonomy, as presented in Table 2, which categorizes the research along the axes of *Core Task Objective*, *DRL Method/Architecture*, and *Key Enhancement/Approach*.

Table 2. A three-level taxonomy of deep reinforcement learning studies for autonomous navigation and path planning, organized by core task objective, algorithmic architecture, and key enhancement or approach.

Core Task Objective	**DRL Method / Architecture**	**Key Enhancement / Approach**	**Sources**
Mapless Navigation	Value-Based (DQN)	Discrete Action Space	[28], [28]

Core Task Objective	DRL Method / Architecture	Key Enhancement / Approach	Sources
	Policy-Based / Actor-Critic	Continuous Control	[29], [30]
		Curiosity-Driven Exploration	[31]
		Imitation Learning / Reinforced Imitation	[32]
		Hybrid Aerial-Underwater Vehicle	[33]
	Memory-Based	Limited Environment Knowledge	[34]
Map-Based / Path Planning	Value-Based (DQN)	Complicated Environments	[35], [36]
		Dynamic Path Planning	[37]
		Mobile Service Robot	[38]
	Actor-Critic (e.g., PPO, A2C)	Map-Based DRL	[39]
		Reward Shaping & Generalization	[40]
		Adaptive Replanning Strategy	[41]
		Voronoi-based Multi-Robot	[42]
		Successor Features for Generalization	[43]
	Fuzzy Logic Integration	Improved Training Convergence	[44]
Autonomous Exploration	Frontier-Based	Benchmarking (Explore-Bench)	[45]
	Goal-Driven	Goal-driven deep exploration	[46]
		Cognitive Exploration	[47]
	Graph-Based	Exploration under Uncertainty	[48]
		Large-Scale Exploration	[49]
		Self-learning Exploration & Mapping	[50]

Core Task Objective	DRL Method / Architecture	Key Enhancement / Approach	Sources
		Risk-Sensitive Exploration	[51]
Obstacle Avoidance / Local Planning	Behavior-Based Methods	A Behavior-based mobile robot navigation method	[52]
	Vision-Based (DQN/CNN)	Vision-based obstacle avoidance	[53]
		UAV Obstacle Avoidance & Target Tracking	[54]
		Semantic Navigation in Dynamic Environments	[55]
	Safe Local Planning	Rough Terrain / Uneven Outdoor	[56], [57]
		Deployment into Conventional Systems (Arena-Rosnav)	[58]
Sim-to-Real Transfer & Deployment	Domain Randomization / Transfer	Cluttered Rough Terrain Pipeline	[59]
		Virtual-to-Real Continuous Control	[29]
		Reward Shaping for Generalization	[40]
	Framework/Tools	Air Learning Gym (Aerial)	[60]
		Pic4rl-gym (ROS2 Framework)	[61]
		Arena-Rosnav (Deployment into NAS)	[58]
	Incremental / Self-Supervised Learning	Incremental learning for navigation	[62]
		Self-supervised with generalized computation graphs	[63]
Specific Robotic Platforms	Wheeled Mobile Robots	Indoor Navigation	[64], [65], [66]
		Omnidirectional Robot	[67]
		Service Robot Path Planning	[38]

Core Task Objective	DRL Method / Architecture	Key Enhancement / Approach	Sources
		Vineyard Navigation (Position-agnostic)	[68]
		Unmanned Vehicle	[69]
	Aerial Robots (UAV/Drones)	Map-less flying robot	[70]
		Drone navigation using sensor data	[71]
		Quadrotor landing	[72]
		Dynamic environment replanning	[73]
Multi-Robot / Cooperative Systems	Cooperative DRL	Multi-robot collision avoidance	[74]
	Voronoi-Based	Multi-robot autonomous exploration	[42]
Knowledge Transfer & Optimization	Reward Optimization	Reward function optimization & knowledge transfer	[75]
	Genetic Algorithms	Genetic DRL for mapless navigation	[28]
	Landmark Generators	Landmark-aided navigation	[76]
	Target-Driven Visual	Target-driven visual navigation	[4]
	Meta-Learning / Generalization	Successor Features	[43]
	Benchmarking & Evaluation	Benchmarking techniques for autonomous navigation	[77]
		Explore-Bench metrics & datasets	[45]
		Air Learning Gym	[60]
		Pic4rl-gym ROS2 Framework	[61]

Within the *Mapless Navigation* category, a prominent direction seeks to remove the dependence on pre-built metric maps, thereby enabling robots to navigate in unfamiliar, unstructured, or dynamic environments solely through reactive sensor-based policies. The choice of action space representation fundamentally shapes the algorithmic approach in this subfield. Several studies adopt a discrete action space, leveraging the classic Deep Q-Network (DQN) algorithm to learn policies that select from a finite set of movement primitives such as turn left, move forward, or turn right. For instance, [28] and [28]*) propose discrete DRL frameworks for mapless navigation, where the agent's policy directly maps raw laser scan or depth image inputs to discrete motor commands. The approach presented in [28]*) is particularly notable as it explicitly compares the performance of discrete action space DQN against continuous control methods, concluding that for the specific task of mapless collision avoidance in cluttered static environments, a carefully tuned discrete policy can achieve comparable success rates while offering simpler training dynamics and greater interpretability [28]*). Complementing this, [28]*) introduces a genetic algorithm to optimize the hyperparameters of the DQN architecture, demonstrating that automatic hyperparameter tuning can significantly improve the sample efficiency and final performance of discrete mapless navigation policies, reducing the need for extensive manual tuning by the researcher [28]*).

However, many robotics tasks necessitate smooth, continuous control signals suitable for differential drive or omnidirectional mobile bases. To address this, a substantial body of work employs actor-critic architectures, particularly Deep Deterministic Policy Gradients (DDPG) and Soft Actor-Critic (SAC), to learn continuous motion policies. The challenge of transferring a policy learned in simulation to a real robot is a central theme in continuous mapless navigation. [29](# "IS17*) proposes a virtual-to-real DRL framework that explicitly addresses this sim-to-real gap, learning a continuous control policy for a mobile robot in a simulated environment and then successfully transferring it to a physical robot without any real-world fine-tuning. The key to their success is the application of domain randomization during training, where the simulator's parameters—such as sensor noise, floor friction, and obstacle placement—are randomly varied to force the policy to learn robust, invariant features [29]*). Similarly, [30]*) also tackles the map-less goal-driven navigation problem with a DDPG-based policy trained entirely in simulation, albeit with the distinction that their approach is evaluated on a real robot after training, achieving a high success rate in reaching specified goal positions without relying on an explicit map [30]*).

The persistent challenge of exploration in sparse-reward environments, where the agent receives no feedback until it stumbles upon the goal, has inspired several innovative enhancements to standard DRL for mapless navigation. The work of [31]*) integrates curiosity-driven exploration into a continuous navigation framework, augmenting the extrinsic reward signal from the task with an intrinsic reward that is proportional to the prediction error of a forward dynamics model. This mechanism encourages the robot to visit states where its predictions are inaccurate, effectively driving it to explore novel regions of the state space and thereby accelerating the discovery of the goal [31]*). In contrast, [32]*) addresses the sample efficiency problem through a technique called reinforced imitation, which leverages a small number of human-provided expert demonstrations. The DRL agent is trained not only from its own experience but also through a behavioral cloning loss that encourages it to mimic the expert's actions, dramatically reducing the number of interactions required to learn a viable policy for mapless goal-seeking [32]*). Two additional studies extend mapless navigation to more challenging platforms: [33]*) adapts a continuous DRL framework for a hybrid aerial-underwater vehicle that must navigate through both air and water, learning a policy that can handle the drastically different dynamics and sensor modalities of the two mediums [33]*).

Meanwhile, [34]*) introduces a memory-based architecture, augmenting the policy network with a recurrent neural network or external memory to endow the UAV with a representation of its past trajectory, which is crucial for obstacle avoidance when the agent has only limited, local sensor information [34]*).

Alongside mapless methods, a significant portion of the research focuses on map-based path planning, where the agent has access to a global representation of the environment, such as an occupancy grid or a topological graph. These approaches aim to learn high-level planning strategies that determine a sequence of waypoints or subgoals, which are then executed by a lower-level controller. Value-based methods, particularly DQN variants, are frequently adopted for this task due to their ability to handle discrete action spaces that naturally correspond to moves on a grid or between nodes in a graph. For example, [35]*) and [36]*) both propose DQN-based frameworks for mobile robot path planning in complicated and static environments. The contribution of [35]*) lies in its use of a convolutional neural network to process a local occupancy grid map, outputting a value for each possible movement direction, while [36]*) formulates the problem as a shortest-path finding task where the DQN agent learns to output the optimal next grid cell to visit [35]*), [36]*). A natural extension of this work is the integration of dynamic obstacles. The study by [37]*) proposes a dynamic path planning method using DQN, where the state representation includes the positions and velocities of moving obstacles, allowing the learned policy to predict their future motion and greedily plan an evasive path in real-time [37]*). Similarly, [38]*) demonstrates the application of a DQN-based path planner on a mobile service robot navigating in a domestic environment, illustrating the practicality of these methods for real-world service scenarios [38]*).

Actor-critic methods have also been extensively applied to map-based planning, offering the advantage of learning continuous policies that can output smooth acceleration or velocity commands rather than discrete grid moves. The study by [39]*) explicitly presents a map-based DRL framework where the policy takes as input a global occupancy grid and a local goal, and outputs continuous linear and angular velocities, demonstrating more fluid motion compared to a discrete planner [39]*). A critical challenge in map-based DRL is generalization across environments with different layouts. [40]*) addresses this by proposing a reward shaping technique that augments the sparse goal-reaching reward with a dense potential-based reward derived from a global planner. This shaped reward provides a smoother learning signal and, crucially, the authors demonstrate that the learned policy generalizes significantly better to unseen map layouts compared to a baseline without reward shaping [40]*). A more sophisticated planning meta-controller is introduced by [41]*), which frames the decision of when to replan a global path as a DRL problem. Instead of using a fixed replanning frequency, the agent learns an adaptive replanning strategy based on the perceived changes in the environment, leading to more efficient navigation in dynamic settings [41]*). The incorporation of fuzzy logic into the DRL training loop is a novel approach proposed by [44]*), where the Q-value estimation is fuzzified to handle uncertainty in the state perception, resulting in improved training convergence and more robust navigation in the presence of sensor noise [44]*).

Autonomous exploration, where the robot must systematically traverse an unknown environment to build a map, is a distinct but closely related problem. This task is inherently uncertainty-driven, as the robot's objective is not simply to reach a goal but to maximize information gain about the world. A benchmark study by [45]*) provides a comprehensive evaluation framework, including datasets, metrics, and baseline implementations, for both frontier-based and DRL-based exploration methods, establishing a standard for comparing future work [45]*). Goal-driven exploration, as exemplified by [46]*), frames exploration as a single agent learning to select navigational goals that lead to the discovery of new areas, effectively learning an exploration policy that is separate from the goal-reaching policy [46]. Taking this a step further, [47]*) proposes a framework for cognitive exploration, where the robot explicitly models its own uncertainty and curiosity to guide its exploration strategy, echoing the intrinsic motivation ideas seen in [31]*) but applied to the specific context of building a complete map of the environment [47]*).

Graph-based representations of the environment offer a natural way to model connectivity and space for exploration. The work of [48]*) formulates exploration under uncertainty as a DRL problem on a graph, where the agent's action space corresponds to traversing edges to frontier nodes, and the reward is a function of the reduction in map entropy achieved by exploring each new node [48]*). [49]*) extends this graph-based DRL paradigm to large-scale exploration tasks, demonstrating that the policy can learn to efficiently allocate exploration resources by leveraging a graph representation that abstracts away low-level geometric details [49]*). A self-supervised learning approach is presented by [50]*), where the robot first learns to explore and map using a purely intrinsic reward signal (based on novelty or prediction error), without any external goal specification, and then uses the learned exploration policy to bootstrap a goal-directed navigation system [50]*). Recognizing the potential risks of exploration, [51]*) introduces a risk-sensitive DRL formulation for autonomous exploration. Instead of maximizing the expected information gain, the policy is trained to maximize a risk-averse objective, such as the conditional value-at-risk (CVaR) of the information gain, thereby ensuring that the agent avoids potentially catastrophic decisions, particularly in hazardous environments [51]*).

Obstacle avoidance and local planning, the task of reacting to immediate sensory percepts to avoid collisions while making progress toward a goal, represent a third major research cluster. Early work in this vein adopts a behavior-based approach, as seen in [52]*), which proposes a DRL method that learns to arbitrate between simple avoidance behaviors such as turn left, turn right, or slow down, effectively learning a high-level behavior selector rather than a direct motor controller [52]*). Vision-based obstacle avoidance has gained significant traction due to the low cost and rich information provided by cameras. [53]*) presents a framework where a DQN agent processes raw depth images from a forward-facing camera to output discrete turning commands for a mobile robot, achieving effective collision avoidance in indoor cluttered environments [53]*). In the domain of aerial robotics, [54]*) proposes a DRL method for autonomous UAV obstacle avoidance and target tracking, learning a policy that simultaneously avoids obstacles and maintains visual contact with a moving target, a challenging control problem requiring precise coordination [54]*). For more complex dynamic environments, [55]*) introduces a semantic navigation approach that combines DRL with semantic segmentation. The policy not only perceives depth and occupancy but also understands the semantic categories of objects (e.g., a chair vs. a wall), learning that certain* *obstacles, like a sofa, can be safely approached, while others, like a glass table, must be handled with extreme caution [55]*).

Safe local planning in rough, uneven terrain presents a unique set of challenges, as the robot must not only avoid obstacles but also maintain stability and avoid traversing surfaces that could cause it to tip over or get stuck. The study by [56]*) proposes a DRL method for safe local planning of a ground vehicle in unknown rough terrain, learning a policy that takes as input a terrain heightmap and outputs a safe velocity command, explicitly penalizing terrain traversals that exceed a learned stability threshold [56]*). Similarly, [57]*) introduces a method called TERP (Terrain Reliable Planning), which uses DRL to learn a reliable planner for uneven outdoor environments, focusing on recovering from challenging situations such as slipping on loose gravel or being pushed by branches, demonstrating high robustness in field tests [57]*). The practical deployment of such learned controllers often requires integration into established navigation stacks. This is precisely the motivation behind [58].

A critical and growing research stream focuses explicitly on the *Sim-to-Real Transfer and Deployment* of navigation policies, recognizing that a policy that works flawlessly in simulation is of limited value unless it can be reliably deployed on physical hardware. The study by [59]*) develops a complete sim-to-real pipeline for autonomous robot navigation in cluttered, rough terrain. Their approach combines domain randomization during training, systematic calibration of the real robot's sensor noise, and a lightweight post-processing filter to reduce the sim-to-real gap, culminating in successful deployment of a DRL policy on a physical quadruped robot navigating over rocks and debris [59]*). As previously discussed, [29]*) and [40]*) each offer their own approaches to sim-to-real transfer, with [40]*) emphasizing reward shaping to improve generalization [29]*), [40]*). The creation of specialized simulation environments and software frameworks is a vital enabler of this research. [60]*) presents Air Learning, a high-fidelity gym environment built on top of the AirSim plugin for Unreal Engine, specifically designed for autonomous aerial robot visual navigation. It provides realistic sensor models and physics, making it a suitable testbed for developing and evaluating sim-to-real transfer techniques for UAVs [60]*). In a complementary vein, [61]*) introduces pic4rl-gym, a modular ROS2-based framework for robot autonomous navigation with DRL. Its modularity allows researchers to easily swap out policy architectures, sensor configurations, or reward functions, accelerating the iterative cycle of training and evaluation [61]*). Another important aspect of deployment is the ability to adapt to changing environments. [62]*) proposes an incremental learning approach for autonomous mobile robot navigation, where the DRL policy is continuously fine-tuned online as the robot is deployed in its target environment, allowing it to adapt to gradual changes in lighting, obstacle placement, or floor conditions without forgetting previously learned behaviors [62]*). Similarly, [63]*) presents a self-supervised DRL method based on generalized computation graphs, which enables the robot to learn new navigation skills in a scalable manner by composing and reusing previously learned components, without the need for extensive human supervision or explicit sim-to-real transfer procedures [63]*).

The diversity of *Specific Robotic Platforms* employed in these studies further highlights the breadth of the navigation and path planning research. A substantial number of works focus on wheeled mobile robots operating in indoor environments. For instance, [64]*) demonstrates a DRL approach for real autonomous mobile robot navigation in indoor environments, validating their policy on a physical robot navigating through hallways and rooms [64]*). [65]*) and [66]*) also present novel DRL-based methods for indoor mobile robot navigation, the former emphasizing a comprehensive survey architecture and the latter proposing a complete novel navigation system [65], [66]*). Several works target specific wheeled platforms: an omnidirectional robot is the subject of [67]*), which leverages the robot's holonomic capabilities to learn complex maneuvering behaviors in tight spaces [67]*); mobile service robots are the focus of [38]*); and a unique application to agricultural robotics is presented by [68], which demonstrates position-agnostic navigation in vineyards using DRL, enabling a robot to traverse rows of vines without relying on GPS or global localization [68]*). Ground unmanned vehicles are also studied, as in [69]*), which provides a general framework for autonomous navigation of unmanned vehicles using DRL [69]*).

In the aerial domain, [70]*) was an early work demonstrating the feasibility of using DQN for map-less navigation of a flying robot, learning a policy to avoid obstacles and navigate toward a goal using only on-board sensor data [70]*). A more recent and comprehensive study by [71]*) addresses drone navigation using multiple sensor modalities, fusing camera images with depth and inertial data to learn a robust visual navigation policy for outdoor flight [71]*). High-precision landing is a critical skill for autonomous quadrotors, and [72]*) proposes a DRL approach for autonomous quadrotor landing on a moving platform, learning a policy that can handle the relative motion dynamics and visual servoing required for a successful touchdown [72]*). The challenge of navigating in dynamic environments, where obstacles such as other aircraft or birds may appear unpredictably, is tackled by [73]*) in the context of real-time path planning for a UAV, using a DRL method to determine when and how to replan the trajectory to avoid collisions while maintaining mission progress [73]*).

Finally, a cluster of studies investigates *Multi-Robot / Cooperative Systems* for navigation and exploration. The work of [74]*) proposes cooperative deep reinforcement learning policies for autonomous navigation in complex environments, where multiple mobile service robots must learn to share space and coordinate their movements to avoid collisions while each pursuing their own individual goals. Their framework leverages a centralized training, decentralized execution (CTDE) architecture to learn cooperative behaviors [74]*). [42]*) addresses the problem of multi-robot autonomous exploration using a Voronoi-based DRL approach. In this method, the exploration space is partitioned into Voronoi cells, and each robot learns a policy to optimally explore its assigned cell while also dynamically adjusting its boundary based on the exploration progress of its neighbors, resulting in efficient and coordinated exploration without centralized communication [42]*).

Several additional studies fall into the categories of *Knowledge Transfer & Optimization* and *Benchmarking & Evaluation*. For example, [75]*) explicitly addresses the problem of optimizing the reward function for mobile robot navigation and uses transfer learning to adapt a policy learned in one environment to another, thereby improving sample efficiency [75]*). The integration of genetic algorithms for hyperparameter optimization in DRL for mapless navigation has already been mentioned as [28]*). The use of landmark generators is proposed* *by [76]*), where a separate generative network is trained to output intermediate subgoals or landmarks that guide the DRL navigation policy, effectively breaking down the long-horizon navigation task into a series of shorter subproblems [76]*). Target-driven visual navigation is the focus of [4]*), a seminal work that demonstrates how a DRL agent can be trained to navigate to a specific target object in an indoor scene by learning a visual representation that is invariant to the agent's starting pose [4]*). A meta-learning perspective is adopted by [43]*), which uses successor features to enable a DRL agent to rapidly adapt its navigation policy to new environments or goals by learning a representation of the environment's environmental dynamics that can be recomposed for different tasks [43]*). Finally, comprehensive benchmarking efforts have been essential for standardizing evaluation in this field. The study by [77]*) provides a thorough benchmark of multiple DRL techniques for autonomous navigation, comparing DQN, DDPG, PPO, and SAC on a set of standardized navigation tasks in both simulated and real-world environments, offering valuable insights into the relative strengths and weaknesses of each algorithm for this specific problem class [77]*). As previously noted, [45]*) and [60]*) provide specialized benchmark suites for exploration and aerial navigation, respectively.

Deep Reinforcement Learning for Robotic Manipulation and Physical Interaction

A central ambition of modern robotics is endowing machines with the ability to physically interact with their environment through manipulation, a capability that is foundational for applications ranging from industrial assembly to autonomous surgery. The domain of robotic manipulation presents unique challenges for deep reinforcement learning that distinguish it from the navigation tasks previously discussed. Unlike navigating a mobile robot through open space, manipulation requires precise control of contact forces, the management of complex friction and deformation dynamics, and often demands a high degree of dexterity to handle objects of varying shapes, sizes, and material properties. The consequences of failure are also more immediate and damaging; a collision during navigation may merely stop a robot, but an incorrect grasp or an overly forceful insertion can damage parts, tools, or the robot itself. Consequently, the research within this dimension, comprising **15 studies** as summarized in Table 3, exhibits a distinct emphasis on sample efficiency, sim-to-real transfer strategies, and contact-rich task learning.

Table 3. A taxonomy of deep reinforcement learning studies for robotic manipulation and physical interaction, organized by the core manipulation task, the specific methodology employed, and the key contributions or focus areas.

Core Task	DRL Methodology / Key Focus	Specific Approach / Contribution	Sources
Grasping & Dexterous Manipulation	Foundational & General Methods	Comprehensive survey of DRL for robotic manipulation	[78]
		Asynchronous off-policy updates for robotic manipulation	[79]

Core Task	DRL Methodology / Key Focus	Specific Approach / Contribution	Sources
		Efficient, general, and low-cost dexterous manipulation	[80]
		A comprehensive DRL framework for grasping	[81]
	Task Simplifications & Transfer	Comparing task simplifications for closed-loop picking	[82]
		Real2Sim or Sim2Real policy adaptation for visual insertion	[83]
		Grasping on the Moon from 3D Octree observations	[84]
	In-Hand Manipulation	Dexterous in-hand manipulation of slender objects using tactile sensing	[85]
Task-Specific & Contact-Rich Control	Vision-Based Motion Control	Vision-based DRL for general robotic motion control	[86]
		Visual foresight through model-based DRL	[87]
		Visual inputs and natural rewards for industrial insertion	[88]
	Arm Control & Task Training	Robotic arm control and task training through DRL	[89]
		Learning kinematic feasibility for mobile manipulation	[90]
	Door Opening	Versatile door opening with adaptive position-force control and RL	[91]
	Soft Tissue Manipulation	Manipulating soft tissues for autonomous robotic surgery	[92]

A substantial body of work within this dimension has established foundational methodologies and benchmarks for learning manipulation skills. The comprehensive survey presented by [78] provides an essential overview of the state-of-the-art in DRL for robotic manipulation, synthesizing the algorithmic landscape from early DQN-based discrete action grasping to more advanced continuous control methods like DDPG and SAC. This survey is particularly valuable for its structured discussion of the key challenges that persist in the field, including the sparse reward problem inherent in manipulation tasks where a successful grasp yields a single positive signal after many failed attempts, the difficulty of learning generalizable policies that work across a diverse range of objects, and the critical role of reward function design in shaping learned behaviors [78]. Building upon these foundational concepts, the work by [79] introduces an asynchronous off-policy update mechanism specifically tailored for robotic manipulation. Their approach allows multiple actors to gather experience in parallel in the real world while a single learner processes the collected data asynchronously, thereby decoupling the data collection and learning pipelines. This methodology dramatically accelerates the training process for a robotic arm learning to grasp and lift objects, demonstrating that asynchronous updates are a practical solution to the slow data collection problem that plagues real-world robotics [79].

The pursuit of dexterity, or the ability to manipulate objects with a high degree of freedom, is a hallmark of advanced manipulation research. The study by [80] directly tackles this challenge by proposing a framework for efficient, general, and low-cost dexterous manipulation. Unlike many approaches that rely on expensive multi-fingered robotic hands or specialized simulation environments, this work demonstrates that a DRL policy trained on a simplified simulation of a low-cost gripper can learn complex in-hand rotation behaviors that generalize to the physical hardware. The authors attribute their success to a carefully designed reward function that encourages the agent to explore diverse finger positions and a domain randomization strategy that makes the policy robust to the inevitable discrepancies between the simulated and real gripper [80]. In a complementary vein, [81] provides a comprehensive DRL framework for the general problem of robotic grasping. This work systematically evaluates the performance of different DRL algorithms, including DQN, DDPG, and PPO, on a standardized grasping benchmark, concluding that while DDPG offers superior sample efficiency for continuous control tasks, PPO often yields more stable and repeatable grasps [81].

Several studies explicitly address the practical challenges of learning manipulation policies from scratch. The research by [82] offers a pragmatic perspective by comparing different task simplifications for learning a closed-loop object picking policy. The authors find that simplifying the action space, for instance by constraining the gripper's approach angle or using a pre-grasp alignment behavior, dramatically reduces the number of training episodes required to achieve a high success rate compared to learning a full 6-DoF end-effector policy from the start. This finding is of significant practical importance, suggesting that a careful decomposition of the manipulation problem can be more impactful than algorithmic improvements alone [82]. The sim-to-real transfer challenge, which is particularly acute in manipulation due to the high sensitivity of contact dynamics to simulation inaccuracies, is the central focus of [83]. This work investigates the problem of visual insertion, where a robotic arm must insert a peg into a hole, a classic benchmark for precision assembly. The authors propose a novel Real2Sim policy adaptation strategy, where a policy is initially trained in a high-fidelity simulator and then fine-tuned on the real robot using a small number of real-world demonstrations, effectively bridging the gap between the simulated and real visual appearances and dynamics [83].

The challenge of manipulation in extreme environments is explored by [84], which investigates grasping on the Moon using DRL. This work is notable for its use of 3D Octree observations, a sparse volumetric representation of the environment that is highly memory-efficient and well-suited for the limited computational resources available on a planetary rover. The DRL policy learns to identify and grasp rocks of various shapes from a point cloud input, demonstrating the viability of autonomous manipulation for future lunar missions [84]. Another specialized but deeply insightful study is the work on dexterous in-hand manipulation of slender cylindrical objects by [85]. This task requires the robotic hand to roll a cylindrical object, such as a pen or a screwdriver, between its fingertips without dropping it. The authors demonstrate that integrating tactile sensing—specifically, force and slip sensors embedded in the fingertips—into the DRL policy is essential for success. While a purely visual policy fails to maintain a stable grip due to occlusion and visual latency, the tactile-augmented policy learns to reactively adjust its grip forces in response to incipient slip, achieving a high level of dexterity [85].

Moving beyond general grasping, a significant cluster of research addresses task-specific and contact-rich control problems that demand greater interaction understanding. The work by [86] presents a general framework for vision-based DRL for robotic motion control, including manipulation primitives. Their key insight is that learning a latent representation of the visual state can greatly improve the sample efficiency of downstream policy learning. By pre-training a deep autoencoder on a large corpus of random interaction data, they provide the DRL agent with a compact and meaningful state representation, which then enables the agent to learn a visuomotor policy for reaching and grasping with significantly fewer real-world interactions than a method that learns the representation from scratch [86]. This theme is further developed by [87], which introduces the concept of "visual foresight" through model-based DRL. Instead of learning a direct state-to-action mapping, the agent learns a predictive model of the world—specifically, how images change as a result of its actions. This model is then used to perform planning in the latent image space, effectively allowing the robot to "imagine" the consequences of a sequence of actions before executing them. This model-based approach is shown to be dramatically more sample-efficient than model-free alternatives for several vision-based manipulation tasks, including pushing and pick-and-place operations [87].

Industrial insertion tasks, which require high precision and the ability to handle jamming or misalignment, are addressed by [88]. This work focuses on learning insertion policies using only visual inputs and natural rewards. The natural reward is defined as simple task success or failure—the peg is either fully inserted or it is not—without any engineered shaping terms. The challenge here is the extreme sparsity of the reward, as millions of random trial-and-error interactions may be required to achieve a single successful insertion. To overcome this, the authors employ a combination of curriculum learning, where the robot starts with an easier task (e.g., a larger peg or an aligned socket) and gradually progresses to the target difficulty, and hindsight experience replay, where failures are relabeled as successes for slightly modified goals. This combination enables the agent to learn a high-precision insertion policy from a purely visual input and a binary reward signal [88].

The control of robotic arm itself is the subject of study in [89], which presents a DRL framework for general arm control and task training. This work is notable for its attempts to learn multiple distinct manipulation tasks—such as reaching, pushing, and grasping—within a single neural network architecture, demonstrating that a shared policy can acquire a varied repertoire of motor skills [89]. The problem of mobile manipulation, where a manipulator arm is mounted on a mobile base, introduces additional kinematic complexity. The research by [90] specifically addresses the challenge of learning kinematic feasibility for such systems. The robot must learn not only how to move its arm to grasp an object but also how to position its mobile base to make the grasp reachable, given that the arm has a limited workspace. The DRL policy learns to reason about the joint feasibility of base placement and arm motion, enabling the robot to grasp objects that would otherwise be out of reach [90].

Finally, two studies address highly specialized and high-stakes manipulation domains. The work by [91] tackles the problem of autonomous door opening with a mobile manipulator. Door opening is a classic contact-rich manipulation task that requires precise force control to unlatch the handle while simultaneously moving the base to open the door, all while managing the interaction forces to avoid damage. The authors propose a hybrid method that combines adaptive position-force control with DRL. The DRL policy learns the high-level strategy for unlatching and pushing the door, while a low-level force controller ensures that the applied forces remain within safe limits, demonstrating a successful synergy between classical control and modern learning [91]. Perhaps the most safety-critical manipulation application is autonomous robotic surgery, which is the focus of [92]. Manipulating soft tissues, such as lifting a flap of skin or retracting an organ, is extremely challenging due to the non-linear, deformable, and often fragile nature of the tissue. The DRL agent must learn to apply just enough force to lift the tissue without tearing it, a constraint that is both a performance metric and a safety requirement. The authors demonstrate that a DRL policy can learn to manipulate soft tissues in a simulated surgical environment, and they validate the feasibility of transferring the learned policy to a physical surgical robot, paving the way for greater autonomy in surgical procedures [92].

Multi-Agent Systems and Swarm Intelligence

The extension of deep reinforcement learning from single-agent settings to environments inhabited by multiple interacting agents introduces a profound increase in complexity while simultaneously unlocking the potential for emergent collective behaviors that far exceed the sum of individual capabilities. In these scenarios, agents must not only learn a policy for their own task but also account for the actions, policies, and learning dynamics of other agents, a phenomenon that can lead to non-stationary environments, credit assignment challenges, and coordination dilemmas that are absent in single-agent formulations. The included studies within this dimension, totaling **23 papers**, demonstrate a rich diversity of approaches that address these fundamental challenges across a spectrum of coordination strategies and application domains. We structure our analysis of this literature according to a taxonomy that captures the primary coordination strategy, the specific application or scenario, and the key technical approach, as presented in Table 4.

Table 4. A taxonomy of deep reinforcement learning studies for multi-agent systems and swarm intelligence, organized by coordination strategy, application scenario, and specific technical approach

Coordination Strategy	**Application / Scenario**	**Specific Approach / Technique**	**Sources**
Centralized / Federated Learning	Task & Role Allocation	Federated Learning for task scheduling in heterogeneous robotic systems	[93]
		Autonomous and adaptive role selection for collaborative area search	[94]
	Communication & Scheduling	Event-triggered communication for distributed cooperative transport	[95]
Decentralized / Distributed	Navigation & Path Planning	Collision Avoidance	End-to-end decentralized navigation in unknown complex environments
		Motion planning among dynamic, decision-making agents	[96]
		Fully distributed collision avoidance for safe and efficient navigation	[97]
		Distributed collision avoidance for navigation in complex scenarios	[98]
		Decentralized non-communicating multiagent collision avoidance	[99]
		Towards optimally decentralized multi-robot collision avoidance	[100]
		Cooperative multi-robot navigation in dynamic environment	[101]
		Decentralized structural-RNN for robot crowd navigation	[102]

Coordination Strategy	Application / Scenario	Specific Approach / Technique	Sources
		Multi-robot path planning based on a DQN algorithm	[103]
	Exploration & Information Gathering	Multi-robot information gathering with deep reinforcement learning	[104]
		Decentralized multi-robot exploration with macro actions	[105]
Swarm & Flocking	Collective Behavior	Guided deep reinforcement learning for swarm systems	[106]
		Bio-inspired collision avoidance in swarm systems via deep reinforcement learning	[107]
		A deep reinforcement learning strategy for autonomous robot flocking	[108]
Cooperative Tasks	Object Transport	Decentralized control of multi-robot system in cooperative object transportation	[109]
	Encirclement & Pursuit	Decentralized multi-agent pursuit using deep reinforcement learning	[110]
		Multi-robot target encirclement control with collision avoidance	[111]
	Area Search	Autonomous and adaptive role selection for multi-robot collaborative area search	[94]
General Framework	Multi-Agent Control & Decision-Making	Learning data-driven decision-making policies in multi-agent environments	[112]

Coordination Strategy	Application / Scenario	Specific Approach / Technique	Sources
		Cooperative multi-agent control using deep reinforcement learning	[113]
		Multi-agent behavioral control system using deep reinforcement learning	[114]

A primary axis of differentiation in the reviewed literature is the choice between centralized training with decentralized execution (CTDE) and fully decentralized learning. The CTDE paradigm, which is implicitly employed by many shared-reward cooperative settings, allows agents to access global state information during training but forces each agent to act on its local observations during execution. This approach mitigates the non-stationarity problem during training while maintaining the scalability of decentralized execution. The fully decentralized approach, in contrast, makes no assumptions about communication or shared learned models, and each agent must learn its policy purely from its own local experience. The works on collision avoidance, which constitute a significant proportion of this dimension's literature, frequently adopt a fully decentralized framework due to its scalability and robustness to communication failures.

The challenge of decentralized, non-communicating multi-agent collision avoidance is addressed by several studies that represent a progression toward increasingly sophisticated coordination. The work presented in [100] proposes a framework for optimally decentralized multi-robot collision avoidance, where each robot independently learns a local policy that aims to avoid collisions while making progress toward its own goal. The key insight of this work is that the optimal local policy can be learned by treating other robots as part of a dynamic environment, effectively learning a reactive collision avoidance behavior that generalizes to varying numbers of neighboring agents [100]. This foundational approach is extended by [115] which proposes an end-to-end decentralized method for navigation in unknown complex environments. Instead of using hand-crafted observation features, the policy directly maps raw sensor data, such as laser scans or depth images, to continuous velocity commands, learning to implicitly model the presence and intentions of other agents from these sensor streams [115].

A more holistic approach to multi-robot collision avoidance is presented by [97] and [98], both of which propose fully distributed frameworks for safe and efficient navigation. The method in [97] addresses the problem of congestion, where multiple robots simultaneously converge on narrow passages or intersections. Their distributed policy learns to yield right-of-way to other robots based on their relative proximity to the intersection, effectively learning a decentralized traffic management protocol without explicit communication [97]. Similarly, [98] focuses on the complexity of navigation in scenarios with multiple decision-making agents, such as human pedestrians and other robots, and learns a policy that is robust to the unpredictable motions of these dynamic entities [98]. The study by [99] specifically targets the most challenging scenario: decentralized collision avoidance without any form of explicit communication. The agents in this framework must infer the intentions of other agents purely from their observed trajectories, learning to anticipate their future motions and plan their own path accordingly. This is a significant step toward practical scalability, as it removes the dependency on any communication infrastructure [99].

The problem of robot crowd navigation, where a single robot must navigate through a densely packed group of human pedestrians, is tackled by [102] using a decentralized structural-RNN architecture. Unlike the symmetric agent-agent interactions considered in the multi-robot collision avoidance works, crowd navigation introduces an asymmetry, as pedestrians have their own objectives and are not contributing to a shared goal. The structural-RNN encodes the spatial relationships between the robot and each pedestrian, modeling the influence of each pedestrian on the robot's optimal action through a learned interaction graph. This approach allows the policy to handle crowds of varying densities and to generate socially compliant behavior, such as passing on the right or maintaining a comfortable distance from pedestrians [102].

For scenarios that involve more than just collision avoidance, but are related to, simple collision avoidance, the work by [103] applies the DQN algorithm to multi-robot path planning. This method formulates the mission to a graph-based representation of the environment and learns to select the next node for each robot to visit, using a centralized coordination mechanism to resolve conflicts. While effective, this approach requires global knowledge of all robot positions and their planned paths, which may not scale to very large robot teams [103]. The work by [101] extends the scope to cooperative multi-robot navigation in dynamic environments, where robots must learn to not only avoid collisions with each other but also with moving obstacles such as doors or shelves that are being moved by humans or other robots [101].

A distinct but related application is cooperative target encirclement and pursuit. The study by [110] proposes a decentralized multi-agent pursuit algorithm where a team of pursuers must learn to coordinate their movements to capture a single evader. The agents learn a distributed policy that assigns each pursuer a role, such as blocking an escape route or closing in from a specific direction, entirely through the shared reward signal of capturing the evader. This emergent role allocation, without explicit communication, is a powerful demonstration of decentralized cooperative learning [110]. The work by [111] addresses a more structured encirclement problem, where a team of robots must surround and maintain a formation around a target object without colliding with each other. Their DRL framework learns a control policy that combines a repulsive potential field for collision avoidance with an attractive potential toward a dynamically computed encirclement point, with the DRL policy learning to weigh these competing objectives optimally [111].

The challenge of cooperative object transport is addressed by [95] and [109]. The work in [95] introduces an event-triggered communication mechanism for a team of robots that must coordinate to transport a large, heavy object to a specified goal location. Instead of communicating continuously, each robot transmits its state only when a specific event occurs, such as a deviation from the planned path or a perceived change in the object's dynamics. This event-triggered scheme drastically reduces the communication bandwidth required while maintaining the coordination necessary for successful transport [95]. In contrast, [109] presents a fully decentralized control framework for cooperative object transportation. In their setup, each robot grasps the object and must learn to apply forces that are consistent with the actions of its teammates. The DRL policy learns to infer the intention of other agents from the force feedback it receives through the object, effectively using the object itself as a communication channel to coordinate its actions [109].

The exploration of unknown environments by a team of robots is a canonical multi-agent problem that directly benefits from DRL. The work by [104], abbreviated as DeepIG, proposes a DRL method for multi-robot information gathering. The objective of the team is to maximize the total information collected about an unknown scalar field, such as a temperature gradient or a contamination plume. Each robot learns to balance exploring new regions against exploiting areas with high expected information density, and the team-level reward is structured to discourage redundant sampling, thereby forcing the robots to spread out and cover more area [104]. The study by [105] addresses the problem of decentralized multi-robot exploration by introducing the concept of macro-actions. Instead of learning a policy that outputs low-level motor commands, the DRL agent learns to select from a set of high-level behaviors, such as "explore frontier cell" or "return to base." This hierarchical structure reduces the complexity of the exploration problem, allowing the DRL policy to focus on the long-term strategic aspects of coordination rather than on low-level motion control [105].

The coordination strategy of role-based control is explicitly addressed by [94] in the context of multi-robot collaborative area search. In this scenario, a heterogeneous team of robots, each with different sensing capabilities (e.g., a camera-equipped robot and a thermal sensor robot), must search an area for a target. The DRL policy learns to dynamically and autonomously select which role each robot should assume at any given time, such as "searcher," "tracker," or "communicator," based on the current state of the search. This adaptive role selection is crucial for maximizing the team's overall search efficiency, as the optimal role assignment changes as the search progresses and new information is gathered [94].

Federated learning is a technique that is naturally suited to multi-robot systems where communication constraints or data privacy concerns prevent the sharing of raw sensor data. The study by [93] introduces a federated deep reinforcement learning approach for task scheduling in a heterogeneous autonomous robotic system. In their framework, each robot trains a local DRL policy using its own local data and periodically shares only the policy gradients or model parameters with a central server, which aggregates them into a global model. This approach allows the robots to benefit from collective experience without explicitly sharing their raw sensor observations, which is particularly advantageous in large-scale or security-sensitive deployments [93].

The pursuit of emergent collective behaviors in swarm systems is captured by a cluster of papers focusing on flocking and swarm guidance. The study by [108] presents a DRL strategy for achieving autonomous robot flocking, a behavior inspired by the coordinated motion of flocks of birds. Each robot in the swarm learns a local policy that balances three fundamental Reynolds flocking rules: separation (avoid crowding neighbors), alignment (steer toward the average heading of neighbors), and cohesion (steer toward the average position of neighbors).

The DRL policy learns the optimal weighting of these rules in response to the local density and topology of the swarm, generating realistic and robust flocking behavior that is resilient to the addition or removal of individual agents [108].

A more explicitly bio-inspired approach is taken by [107], which proposes a collision avoidance mechanism for swarm systems that mimics the behavior of schools of fish. The DRL policy learns to react to the optical flow patterns generated by neighboring agents, allowing the robot to avoid collisions without explicit distance measurements. This method is particularly appealing for swarms operating in environments where distance sensors are unreliable or unavailable, such as in underwater or low-visibility conditions [107]. The study by [106] addresses the challenge of guiding a large swarm of robots toward a desired global configuration without centralized control. Their method, termed guided deep reinforcement learning, uses a small number of "guide" robots that are controlled by a DRL policy, while the rest of the swarm follows a simple, pre-programmed attraction-repulsion rule. The DRL policy learns to position the guide robots optimally to steer the entire swarm toward a target formation, demonstrating a hybrid approach that combines the learning capabilities of DRL with the scalability of simple, distributed control laws [106].

Finally, the dimension of general multi-agent control and decision-making is addressed by several studies that propose broader frameworks. The work by [112] presents a general framework for learning data-driven decision-making policies in multi-agent environments, focusing on the challenge of learning from limited data and generalizing to unseen agents. Their framework incorporates a mechanism for estimating the policies of other agents from their observed behavior, effectively creating a model of the opponent that helps stabilize the learning process [112]. The study by [113] proposes a more general framework for cooperative multi-agent control, which is evaluated on a series of benchmark coordination tasks including cooperative navigation, predator-prey, and physical deception. Their method is based on a centralized critic that estimates a joint action-value function during training, which is then decomposed into per-agent advantage functions to guide the decentralized policy updates [113]. The work by [114] proposes a multi-agent behavioral control system that integrates DRL with a behavior tree structure. In this framework, the high-level behavior selection is handled by a behavior tree, while the low-level execution of each behavior is learned via DRL. This hierarchical and modular approach provides a degree of interpretability and modularity that is often lacking in monolithic DRL policies, allowing human designers to easily specify safety constraints or ethical guidelines at the behavior tree level while allowing the DRL agent to optimize the execution of each chosen behavior [114].

Safety, Verification, and Robustness

The deployment of deep reinforcement learning in real-world autonomous systems introduces an imperative that extends beyond mere task performance: the learned policies must operate safely, reliably, and predictably under a wide range of conditions, including those not encountered during training. Safety in this context encompasses the avoidance of catastrophic failures, the maintenance of operations within specified constraints, and the assurance of robustness against perturbations, whether adversarial or environmental. Verification, the formal process of proving that a policy will not violate safety specifications, is a complementary but distinct challenge, as the neural network policies produced by DRL are typically opaque and highly non-linear. The 19 studies identified within this dimension exhibit a pluralistic approach, ranging from formal verification methods and probabilistic model checking to practical safety frameworks that incorporate human oversight, physics-informed constraints, and adversarial robustness analysis. We organize this diverse body of work according to a three-level taxonomy, as presented in Table 5, which classifies the studies by their primary approach, sub-approach, and the specific technique or methodology they employ.

Table 5 A structured taxonomy of deep reinforcement learning studies on safety, verification, and robustness, organized by primary approach, sub-approach, and specific technique or methodology

Approach	**Sub-Approach**	**Specific Technique / Method**	**Sources**
Formal Verification & Assurance	Reachability & Model Checking	Reachability verification for reliability assessment	[116]
		Probabilistic model checking for dependability analysis	[117]
	Formal Verification	Formal verification of neural network controlled autonomous systems	[118]
	Assurance & Guidance	Guidance on ML assurance in autonomous systems (AMLAS)	[119]
		Transfer assurance for ML in autonomous systems	[120]
Safe Learning & Control	Safe Reinforcement Learning	Safe learning in robotics: from learning-based control to safe RL	[121]
		Safe RL with model uncertainty estimates	[122]

Approach	Sub-Approach	Specific Technique / Method	Sources
		Enhanced PPO for safe mobile robot navigation	[123]
		Physics-regulated DRL with invariant embeddings	[124]
		Physics-informed ML framework for safe control	[125]
	Control Barrier Functions & Intermittent Control	Hierarchical multi-agent RL with control barrier functions	[126]
		Opportunistic intermittent control with safety guarantees	[127]
Robustness & Adversarial Analysis	Adversarial Attacks	Characterizing attacks on DRL	[128]
		Enhanced adversarial strategically-timed attacks	[129]
	Safety Validation & Testing	Dense RL for safety validation of autonomous vehicles	[130]
		Accelerated sim-to-real RL for collision avoidance	[131]
Training Strategies for Safety	Human-in-the-Loop & Intervention	Efficiently combining human demonstrations and interventions for safe training	[132]
		Learning with training wheels:	[133]

Approach	Sub-Approach	Specific Technique / Method	Sources
		speeding up training with a simple controller	
Dependability Analysis	Dependability Analysis	Dependability analysis of DRL-based robotics and autonomous systems	[117]

Within the domain of formal verification, researchers have made notable strides toward providing guarantees about the behavior of neural network controllers. The studies by [116] and [117] form a complementary pair, both focusing on the dependability of DRL-controlled systems but through different formal lenses. The work by [116] proposes reachability verification as the central technique for reliability assessment. The core idea is to compute the set of all states that a DRL-controlled system can reach from a given initial condition, given the policy's decision boundaries and the system dynamics. If this reachable set does not intersect with any unsafe set (e.g., states representing a collision), then the policy can be certified as safe for that particular initial condition. This method provides a rigorous, sound guarantee, but its computational cost grows rapidly with the dimensionality of the state space and the complexity of the neural network [116]. Building on this foundation, [117] introduces probabilistic model checking to analyze the dependability of DRL-controlled systems. Unlike reachability verification, which yields a deterministic yes-or-no answer, probabilistic model checking computes the probability of satisfying a given safety specification within a finite horizon. This approach is better suited for systems where safety cannot be guaranteed with absolute certainty due to stochastic dynamics or sensor noise, but where a probabilistic bound on failure probability is acceptable. By modeling the DRL policy and the environment as a Markov decision process, probabilistic model checking provides quantitative insights into the reliability of the system [117].

The formal verification of neural network controlled autonomous systems is further advanced by the work in [118]. This study directly addresses the challenge of verifying properties of a neural network controller, such as a DRL policy, without enumerating all possible inputs. Their method leverages techniques from satisfiability modulo theories (SMT) to encode the behavior of the neural network as a set of logical constraints. By then adding constraints that represent the negation of a safety property, they can prove that no input exists that would cause a violation, or conversely, find a counterexample that demonstrates a safety hazard. This approach is complete and sound for feed-forward neural networks with ReLU activations, but its scalability to the large networks often used in DRL remains a significant computational hurdle [118].

The challenge of providing assurance for machine learning in autonomous systems, particularly when moving from training to deployment, is the focus of two complementary works. The development of the AMLAS (Assurance of Machine Learning for Autonomous Systems) guidance framework, as presented by [119], represents a high-level, process-oriented approach. AMLAS does not prescribe a specific algorithm or verification tool but instead provides a structured methodology for constructing an assurance case for a machine learning component. This framework guides system developers through the stages of requirements definition, data management, model selection, verification, and deployment, ensuring that safety arguments are systematically documented and evidence is collected. While not a formal method in itself, AMLAS provides a crucial practical bridge between the theoretical verification techniques and the real-world engineering processes required for certification [119]. The issue of assurance during the transfer of a machine learning model from simulation to reality is directly addressed by [120] in their work on transfer assurance. The authors argue that the performance of a DRL policy can degrade unpredictably upon deployment in its target environment due to distributional shift. Their framework proposes a systematic evaluation of the policy's performance in a set of carefully designed perturbed simulation environments, in order to estimate its robustness and to build confidence that the policy will not fail when transferred. This approach provides a pragmatic middle ground between full formal verification, which may be computationally intractable, and naive deployment without any safety analysis [120].

A substantial proportion of the literature in this dimension is dedicated to the paradigm of safe learning and control, which seeks to incorporate safety constraints directly into the learning process rather than verifying a completed policy. The seminal work of [121] provides a comprehensive survey that positions safe reinforcement learning within the broader context of learning-based control. The authors distinguish between two primary categories of safe RL: methods that modify the exploration process to avoid unsafe states during training, and methods that optimize a constrained objective function, such as maximizing reward subject to a constraint on the expected cost of safety violations. The study provides a foundational taxonomy of these approaches, including constrained MDPs, risk-sensitive criteria, and teacher-student frameworks, that continues to structure the subfield [121].

One of the persistent challenges in safe RL is managing the uncertainty inherent in learned models. The study by [122] directly tackles this problem by proposing a safe RL framework that explicitly incorporates model uncertainty estimates. Their approach learns a probabilistic dynamics model of the environment, which provides not only a prediction of the next state but also a measure of the model's epistemic uncertainty. The DRL policy is then trained to avoid state-action pairs with high predictive uncertainty, as those are regions where the model is likely to be inaccurate and where unanticipated unsafe events could occur. This Bayesian perspective offers a principled way to balance exploration for learning with the caution necessary for safe operation [122].

Practical methods for safe mobile robot navigation are the focus of [123], which proposes an enhanced version of the Proximal Policy Optimization (PPO) algorithm. Their key contribution is to augment the standard PPO objective with a safety penalty term that increases quadratically with the proximity of the robot to obstacles. This constrains the policy to learn trajectories that maintain a safe distance from obstacles, even if taking a riskier, closer path would yield a higher task reward. The resulting policy is demonstrated to have a significantly lower collision rate during both training and testing compared to the standard PPO baseline, validating the efficacy of incorporating a shaped safety penalty into the DRL objective [123].

A more principled approach to constraining learning is provided by the integration of physics-based knowledge. The physics-regulated deep reinforcement learning framework proposed by [124] achieves this by learning invariant embeddings. Instead of simply adding a penalty for violating a constraint, this method learns to map the state space into a latent representation where the dynamics of the system follow known physical invariants, such as conservation of energy or momentum. The DRL policy is then trained on this grounded latent space, which inherently biases the policy toward physically plausible and safe actions. This approach is particularly powerful because it provides a form of safety by construction, rather than through a learned penalty that could be circumvented by the optimizing agent [124]. Similarly, the physics-informed machine learning framework presented by [125] integrates a surrogate model of the system's dynamics, derived from first principles, into the DRL training loop. The surrogate model provides a differentiable approximation of the true dynamics, which is used to guide the policy's optimization toward regions of the state space that are both high-reward and consistent with the physics-based safety constraints. The authors demonstrate the efficacy of their method on a benchmark autonomous driving lane-keeping task, where the learned policy avoids the unsafe oscillations that a purely data-driven DRL policy exhibits [125].

Control barrier functions (CBFs) offer a rigorous, controller-theoretic method for guaranteeing safety, and their integration with DRL is a promising direction. The work presented in [126] proposes a hierarchical multi-agent RL framework with control barrier functions. In this architecture, a high-level DRL policy generates a desired control input, while a low-level CBF controller filters this input to ensure that it does not lead to a violation of safety constraints, such as collisions with other agents or obstacles. The CBF provides a certificate of safety for a candidate control input, modifying it minimally if it would lead to an unsafe state, while preserving it if it is safe. This separation of concerns allows the DRL agent to focus on optimizing task performance, while the formal guarantees of safety are provided by the CBF layer, which operates on a known, tractable model of the system's dynamics [126]. A different but equally compelling control architecture is proposed by [127] in the work on opportunistic intermittent control with safety guarantees. Instead of continuous control, the system operates in a periodic or event-triggered manner, where the controller is only active for short bursts. Between control actions, the system evolves under its open-loop dynamics. The DRL policy learns to determine the optimal timing for these intermittent control interventions, while a separate safety layer monitors the system's state and triggers a control intervention if the system drifts too close to an unsafe region. This approach is particularly relevant for energy-constrained autonomous systems, such as drones or underwater vehicles, where continuous actuation is impractical [127].

The robustness of DRL policies to adversarial manipulation is a critical and often underappreciated aspect of safety. The study by [128] provides a foundational characterization of the types of attacks to which DRL policies are susceptible. They distinguish between attacks on the observation space, where an adversary perturbs the sensor readings perceived by the agent, and attacks on the reward signal, where the adversary corrupts the agent's understanding of its own performance. The work systematically demonstrates that a wide range of standard DRL algorithms are highly vulnerable to even small, carefully targeted perturbations of the observations, leading to catastrophic failures in tasks like navigation and balancing [128]. This vulnerability is further elucidated by [129], which proposes enhanced adversarial strategically-timed attacks. Unlike attacks that perturb every observation, this work focuses on attacks that are applied only at critical decision points, such as when the agent must choose between turning left or right at an intersection. By attacking only at these critical junctures, the adversary can induce policy failure with a much lower attack budget, making the detection of the attack much more difficult. This research highlights that even a rarely activated adversary can be devastatingly effective against a DRL-controlled system [129].

Testing and validation of DRL policies in safety-critical domains like autonomous driving presents its own unique challenges. The dense reinforcement learning framework proposed by [130] addresses the problem of discovering safety-critical scenarios for validation purposes. Instead of relying on random sampling of traffic scenarios, the method trains a separate DRL agent to generate challenging and dense traffic conditions for the autonomous vehicle being tested. This adversarial testing agent learns to create situations that are likely to cause a collision or a near-miss, thereby efficiently exploring the edge cases of the control policy's performance. This framework is a powerful tool for identifying weaknesses in a DRL policy before it is deployed in the real world [130]. A parallel approach to safe sim-to-real transfer is provided by [131], which uses human player data to accelerate the transfer of a collision avoidance policy from simulation to a real robot. By incorporating human gameplay data collected in a simulated environment, the method learns a policy that implicitly internalizes the conservative behaviors exhibited by human players to avoid collisions. This approach demonstrates that leveraging human intuition for safety during training can produce policies that transfer to the real world with fewer failures than policies trained purely on self-supervised DRL [131].

A final, pragmatically focused cluster of research addresses the training process itself, proposing strategies to maintain safety while the agent is learning. The work by [132] presents a framework for efficiently combining human demonstrations and interventions for safe training of autonomous systems in real time. During the initial stages of training, the DRL agent observes a human operator performing the task, collecting a set of demonstration trajectories. As training progresses and the agent begins to act, a human supervisor can intervene at any moment, taking control of the system to avoid a collision or other unsafe act. Each intervention provides a corrective demonstration that is incorporated into the agent's replay buffer. This hybrid approach allows the agent to learn a safe policy from the start, without the need for millions of unsafe trial-and-error interactions [132]. The concept of learning with training wheels is proposed by [133] as a complementary strategy. In this approach, a simple, safe hand-coded controller is used to bootstrap the DRL learning process. The DRL agent is given control only when the simple controller is confident that it can maintain safety, and it is forced to relinquish control back to the simple controller if it begins to deviate toward an unsafe state. This structure allows the DRL agent to explore safely, gradually learning to outperform the simple controller while always operating within its safety envelope. Both [132] and [133] represent training strategies that prioritize safety during the learning process itself, which is essential for any DRL application that cannot afford the cost or risk of catastrophic failure during training.

Advanced Learning Paradigms and Optimization

The pursuit of more sample-efficient, generalizable, and scalable deep reinforcement learning methods has led researchers to move beyond the standard monolithic actor-critic or value-based frameworks. This section synthesizes the contributions of studies that seek to fundamentally alter the learning paradigm itself, either by decomposing the problem hierarchically, integrating complementary learning frameworks, optimizing the learning dynamics via adaptive algorithms or hybrid optimization, or leveraging inverse learning for better system understanding. These advanced paradigms are critical for closing the persistent gap between the simulated success of standard DRL algorithms and their practical deployment in complex, real-world autonomous systems, where sample budgets are limited, tasks are long-horizon, and deployment conditions may shift.

The 13 studies identified within this dimension can be systematically categorized according to the primary paradigm they advance, as shown in Table 6, which organizes the research into four main groups: Hierarchical and Multi-Level Learning, Hybrid and Integrated Frameworks, Optimization and Adaptive Algorithms, and Inverse Learning and Anomaly Handling.

Table 6. A taxonomy of advanced learning paradigms and optimization studies, organized by paradigm, specific technique, and the application or sub-type addressed.

Paradigm	Specific Technique	Application / Sub-type	Sources
Hierarchical & Multi-Level Learning	Hierarchical Reinforcement Learning	Scalable Control	[134], [135]
	Multi-Objective Optimization	Pareto Optimization	[136]
	Continual Learning	Meta-Policy Adaptation	[135]
Hybrid & [135]			
Hybrid & Integrated Frameworks	Planning + DRL	Tactical Decision Making	[137], [138]
	Language-Guided Learning	LLM-Enabled RL	[139], [140]
	Generative AI + RL	Next-Gen Autonomy	[141]
Optimization & Adaptation Algorithms	Genetic Algorithm for Parameter Optimization	Hyper-parameter Tuning	[142]
	Adaptive Eligibility Traces	Time-Scale Adaptation	[143]
	Bayesian Fusion / Control Priors	Controller Integration	[144]

Paradigm	Specific Technique	Application / Sub-type	Sources
	Model Predictive Control (MPC) + Actor-Critic	Skill Acquisition	[145]
Inverse Learning & Anomaly Handling	Inverse Reinforcement Learning	Anomaly Detection & Correction	[146]

The first paradigm we examine is Hierarchical and Multi-Level Learning. The foundational idea of hierarchical reinforcement learning (HRL) is to decompose a complex, long-horizon task into a hierarchy of subtasks, each learned by its own policy, which can dramatically improve sample efficiency and generalization. The work by [134] explicitly tackles the problem of scalability in autonomous systems through a multi-level control approach using HRL. In this framework, a high-level manager policy learns to select a sequence of subgoals or subroutines from a library, while a low-level worker policy learns to execute the selected subgoal by outputting primitive motor commands. The authors demonstrate that this hierarchical decomposition allows the system to learn complex navigation and manipulation tasks in simulated environments with significantly fewer total environment interactions compared to a monolithic Flat RL baseline, as the high-level policy can operate on a compressed, abstract state space and the low-level policy can reuse learned skills across different high-level goals [134].

A related challenge is the optimization of policies that must simultaneously satisfy multiple, often conflicting objectives, such as maximizing speed while minimizing energy consumption or task completion time. The study on Multi-Objective Deep Reinforcement Learning (MODRL) by [136] addresses this directly by proposing a framework that learns a set of Pareto-optimal policies. Unlike standard RL, which optimizes a scalar reward function, MODRL optimizes a vector of reward signals. The authors introduce a method that learns a single policy network that can output actions optimized for different trade-offs between the objectives, allowing the system to dynamically select the appropriate trade-off based on the current context or a user-specified preference [136]. This work is crucial for autonomous systems that must operate under varying operational constraints or with competing mission goals.

The challenge of adapting to new tasks without forgetting previously learned ones is the subject of continual learning, and the work by [135] directly bridges this with HRL. Their proposed method uses a hierarchical deep reinforcement learning architecture to learn a meta-policy network for new tasks by efficiently composing and re-using previously learned lower-level policy networks. When confronted with a novel task, the high-level meta-policy learns to select and sequence appropriate low-level skills from a growing library, enabling the agent to adapt rapidly. Importantly, the method includes mechanisms, such as elastic weight consolidation, to prevent the catastrophic forgetting of previously learned skills when the lower-level networks are updated to accommodate new tasks, thereby providing a scalable framework for lifelong learning in autonomous systems [135].

The second major paradigm concerns Hybrid and Integrated Frameworks that combine DRL with other approaches to leverageable techniques, such as classical planning and large language models (LLMs). The integration of search-based planning with DRL is a powerful hybrid strategy, as it leverages the broad, near-optimal exploration of search algorithms with the reactive, real-time decision-making of learned policies. The work by [137] presents a framework that integrates search-based planning with DRL for path planning in advanced autonomous vehicles. In their approach, a DRL policy is trained to select the best planning strategy or to refine the output of a search-based planner, thereby biasing the search toward more promising regions of the state space. The authors show that this hybrid system achieves better performance on long-horizon path planning tasks than either the pure search-based planner or the pure DRL policy alone [137].

The tactical decision-making problem in autonomous driving, which involves high-level choices such as whether to change lanes or merge, is the focus of the work by [138]. This study proposes a framework that combines planning and DRL by using a predictive model of the environment's future states, generated by a planning algorithm, as a key input to the DRL agent's state representation. The DRL policy learns to select from a set of tactical maneuvers based not only on the current sensor observations but also on the predicted future trajectories of other vehicles, enabling more informed and safer decision-making than a purely reactive DRL policy [138].

The emergence of large language models (LLMs) has opened a new frontier for integrating symbolic reasoning and task-level knowledge with DRL for low-level control. The study by [139], titled "Integrative AI framework for robotics: LLM-enabled reinforcement learning in object manipulation and task planning," proposes a system where an LLM is used to parse high-level natural language instructions into a symbolic task plan, breaking down the complex instruction into a sequence of subgoals. A DRL policy is then trained to execute each subgoal, such as "pick up the red cube" or "place it on the green platform." The LLM provides the semantic understanding of the task, while the DRL provides the motor control, creating a powerful and flexible framework for instruction-following robots [139].

A variation of this language-guided theme is presented by [140] in the context of resource-constrained autonomous systems, such as UAVs with limited onboard compute. Their method, Efficient Language-Guided Reinforcement Learning, proposes a lightweight DRL architecture that processes quantized or compressed linguistic embeddings to guide the policy. The key innovation is to show that a language-conditioned policy that is suitable for deployment on a microcontroller-class processor can be learned, enabling a drone to understand and execute simple verbal commands like "fly to the landmark" or "land on the pad," thereby integrating language understanding into the autonomous system's control loop without requiring cloud connectivity [140].

A broader vision of integrating reinforcement learning with generative AI models for creating next-generation autonomous systems is presented in the study by [141]. This work discusses a conceptual framework where generative models, such as variational autoencoders or generative adversarial networks, are used to synthesize diverse and challenging training scenarios for the DRL agent, addressing the problem of training data scarcity. Furthermore, it explores the use of generative models to predict high-dimensional future sensor observations, enabling a form of model-based DRL that can plan in the generated observation space. While this study is more forward-looking than some of the other application-focused works, it articulates a clear and promising trajectory for the field [141].

The third paradigm we examine groups together Optimization and Adaptive Algorithms that refine the DRL training process itself, moving beyond standard stochastic gradient descent and fixed hyperparameters. The optimization of DRL hyperparameters, such as learning rates, exploration noise, and network architectures, is a notoriously expensive and ad-hoc process. The work presented in [142], titled "Deep reinforcement learning using genetic algorithm for parameter optimization," directly tackles this bottleneck. The authors propose a framework where a genetic algorithm evolves a population of DRL agents, each with different hyperparameter configurations. The performance of each agent on the task is used as its fitness, and genetic operators such as crossover and mutation are applied to generate offspring configurations. The best-performing configuration is then used for final training. This automated approach to hyperparameter tuning significantly reduces human effort and can discover non-intuitive parameter combinations that yield superior performance on a benchmark robotic control task [142].

Another core algorithmic challenge in DRL is the use of eligibility traces, which blend Monte Carlo and temporal-difference learning to control the bias-variance trade-off. The standard approach uses a fixed trace decay parameter λ. The study by [143] proposes a novel method for adaptive and multiple time-scale eligibility traces for online deep reinforcement learning. Instead of a single, global λ, their method learns to control the trace decay rate in a state-dependent and time-dependent manner, effectively allowing the agent to dynamically adjust its bootstrapping strategy. This adaptive approach is shown to be particularly beneficial in robotic problems with a time-varying environment, where the optimal balance between bias and variance shifts as the robot's dynamics change or as the environment's structure is revealed, leading to faster and more stable convergence compared to fixed-trace methods [143].

A different form of optimization is the integration of a learned controller with a prior, often hand-crafted, controller. The work on Bayesian Controller Fusion by [144] presents a principled framework for this. The idea is to fuse the output of a DRL policy with the output of a classical controller, which encodes domain knowledge or safety constraints, using a Bayesian approach. The fusion weight is learned, allowing the system to trust the DRL policy when it is confident and the classical controller when it is not. This provides a soft, probabilistic mechanism for leveraging a control prior, preventing the DRL agent from making fatal errors during early exploration and ensuring graceful performance even in novel situations where the learned policy is uncertain [144].

A final example of algorithmic optimization is the work on Model Predictive Actor-Critic (MPAC) by [145]. This method directly integrates model-predictive control (MPC) with an actor-critic architecture. In standard DRL, the actor learns a policy, and the critic learns a value function. In MPAC, the actor is replaced by a model-predictive controller that uses a learned dynamics model to plan a sequence of actions in a receding horizon fashion. The critic, meanwhile, learns a value function that is used as a terminal cost for the MPC planner. This hybrid approach combines the sample efficiency of model-based planning with the asymptotic performance of model-free RL. The authors demonstrate that MPAC can accelerate robot skill acquisition significantly, enabling a simulated robotic arm to learn a dexterous manipulation task with far fewer interactions than state-of-the-art model-free algorithms [145].

The fourth and final paradigm within this section is Inverse Learning and Anomaly Handling. The study on anomaly detection and correction by [146] proposes a method that uses inverse reinforcement learning (IRL) for optimizing autonomous systems. The core insight is that standard DRL requires a well-engineered reward function, which is often difficult to specify for complex tasks. IRL, in contrast, seeks to infer the reward function from observed expert demonstrations. This study extends IRL to the domain of anomaly detection. By learning a reward function from normal operational data, the system can then detect anomalies as state-action sequences that deviate sharply from the behavior implied by the learned reward. When an anomaly is detected, the system can then re-plan or adjust its behavior to correct it, effectively using the IRL-derived reward as a diagnostic tool. This provides a principled framework for building more robust autonomous systems that can detect and recover from both hardware failures and unexpected environmental conditions [146].

Domain-Specific Applications and Emerging Frontiers

The breadth of deep reinforcement learning research extends far beyond the canonical problems of navigation, manipulation, and multi-agent coordination, finding fertile ground in highly specialized application domains that demand tailored algorithmic adaptations and present unique operational constraints. This dimension of the literature, comprising 75 studies, is the largest in our corpus, reflecting the natural trajectory of a mature technology that is being increasingly applied to solve concrete problems across diverse fields. We note that the sheer volume of studies in this category captures not only the application of DRL to specific robotic platforms but also the emergence of entirely new frontiers, such as the integration of quantum computing with DRL, the development of neurochip-driven edge robotic systems, and the exploration of biologically-inspired learning paradigms like spiking neural networks. To provide a coherent synthesis, we have organized these studies according to a taxonomy that reflects both the domain of application and the novelty of the underlying approach, as presented in Table 7.

Table 7. A taxonomy of domain-specific applications and emerging frontiers in deep reinforcement learning for autonomous systems and robotics, organized by domain and application, algorithmic approach, and key platform or focus.

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
Autonomous Driving & Ground Vehicles	End-to-End & Perception-Driven	Autonomous Driving Systems (General)	[147], [148], [149]
		Urban Autonomous Driving (Latent DRL)	[150]
		Highway Driving	[149]
		Overtaking & Lane Changing	[151]
		Intersection Navigation (Occluded & General)	[152], [153]

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
		Simulation & Validation (CARLA, F1tenth)	[154], [155]
		Distributed DRL on the Cloud	[156]
		Hierarchical & BEV Perception	[157]
		Input-Agnostic & Dynamic	[158]
		Advanced Planning (RL + Inverse RL)	[159]
		Sensor Fusion & Control	[160]
	Racing & High-Speed Control	Champion-level Drone Racing	[161]
		Autonomous Drone Racing	[162]
		High-speed Autonomous Racing (Trajectory-aided)	[163]
		Reward Design & Hyperparameter Tuning for Racing	[164]
Aerial Systems & Drones	Aerial Systems (General)	How to train your robot with DRL (lessons)	[165]
		Perception for Autonomous Systems (PAZ)	[166]
		On-Device Real-Time DRL (FOMO)	[167]
	Vision-Based & Landing	Monocular Vision-based Flight	[168]
		Autonomous Planetary Landing (Transfer Learning)	[169]
		Autonomous Landing on Moving Platform	[170], [171]
		Visual Foresight & Sim-to-Real Flight	[172]

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
	UAVs & Multi-UAV Systems	Multi-UAV Adaptive Path Planning	[173]
		Parallel Distributional Prioritized DRL for UAVs	[174]
		Autonomous Decision-Making for Cooperative UAVs (Search & Rescue)	[175]
		Thermal Soaring (Vulture-Inspired)	[176]
	Motion Control & Navigation	Motion Control in Multi-rotor Aerial Robots	[177]
		Deep RL for Autonomous Robotic Tensegrity (ART)	[178]
		BND*-DDQN for Autonomous Steering	[179]
Marine & Underwater Systems	Underwater & Surface Vessels	AUV Position Tracking Control (End-to-End DRL)	[180]
		Deep Interactive RL for AUV Path Following	[181]
		Deep RL based Tracking Control of an Autonomous Surface Vessel	[182]
		COLREG-compliant Collision Avoidance for USV	[183]
Human-Robot & Social Interaction	Socially-Aware & Shared Autonomy	Shared Autonomy via DRL	[184]
		Socially Aware Motion Planning	[185]
		Crowd-Robot Interaction (Attention-based DRL)	[186]
		Socially Appropriate Robot Approaching Behavior	[187]

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
		Collision Avoidance in Pedestrian-Rich Environments	[188]
	Collaborative & Interaction	Human-centered Collaborative Robots with DRL	[189]
		Cycle-of-Learning for Autonomous Systems from Human Interaction	[190]
		Human-in-the-Loop for Efficient Learning	[191]
		Integrating Expert Guidance for Safe Overtaking	[151]
		Prediction-based Path Planning for Human-Robot Collaboration in Construction	[192]
Legged & Bio-inspired & Micro-Aerial	Legged Robotics	Robust Biped Locomotion (RL on top of Analytical Control)	[193]
		Feedback Control for Cassie with DRL	[194]
		Humanconquad: Human Motion Control of Quadrupedal Robots	[195]
		Automated DRL Environment for Modular Legged Robot	[196]
		Learning to Walk: Spike-based RL for Hexapod CPG	[197]
Medical & Surgical & Inspection	Surgical & Medical Robotics	Deep RL based Semi-autonomous Control for Robotic Surgery	[198]
		Autonomous Navigation of an Ultrasound Probe (Standard Scan Planes)	[199]
		Ultrasound-guided Robotic Navigation with DRL	[200]

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
	Inspection & Infrastructure	Machine Learning for Robotic and Autonomous Inspection	[201]
Industrial & Construction	Industrial & Construction	Autonomous Robotic Additive Manufacturing (Distributed Model-free DRL)	[202]
		Deep RL for Real-time Assembly Planning in Robot-based Construction	[203]
		Wheel Loader Scooping Controller using DRL	[204]
Autonomous Defense & Security	Defense & Security	Reinforcement Learning in Autonomous Defense Systems	[205]
Specialized Emerging Frontiers	Quantum & Spiking Neural Networks	Quantum Deep RL for Robot Navigation	[206]
		Quantum-Enhanced RL for Dynamic Path Planning	[207]
		Exploring Spiking Neural Networks for DRL in Robotic Tasks	[208]
		Learning to Walk: Spike-based RL for Hexapod CPG	[197]
	Neurochip & Edge Systems	Robust Iterative Value Conversion for Neurochip-driven Edge Robots	[209]
	IoT & Embedded Systems	Deep RL for Autonomous Internet of Things	[210]
	Modular & Task Allocation	Modular Task Allocation for Autonomous Mobile Systems	[211]
General & Foundational Applications	Foundational & Theoretical Applications	Applications of Neurocomputing & Stochastic Models	[212], [213]

Domain / Application	DRL Approach / Key Focus	Specific Platform / System	Sources
		Funneled Bayesian Optimization for Design, Tuning and Control	[214]
		Exploring Applications in Real-World Systems (General)	[147], [215], [216]
		Societal Implications of DRL	[217]
		Adaptive Hybrid Algorithms for Real-Time Decision-Making	[191]
		Reinforcement Learning Applications: Traffic to Robotics	[215]
		Deep Learning and RL for Autonomous Unmanned Aerial Systems	[216]
		Reinforcement Learning Behavioral Control for Nonlinear Autonomous System	[218]

A prominent and voluminous sub-dimension within this analysis is the application of DRL to **Autonomous Driving & Ground Vehicles**. This domain has attracted immense research interest, driven by the clear societal and economic impact of self-driving technology. The body of work in this area spans a spectrum from end-to-end perception-to-control systems to more modular, hierarchical architectures that integrate DRL for specific tactical decisions. Early and influential approaches, such as those by [147] and [148], established frameworks for applying DRL to the autonomous driving problem as a whole, demonstrating that a DRL agent could learn to navigate complex urban scenarios in simulation using raw sensor inputs or highly processed state representations. These foundational works set the stage for a wave of research that refined the algorithmic architecture for specific driving sub-tasks. For instance, the problem of highway driving, which involves maintaining a safe speed and following distance while potentially changing lanes, was directly addressed by [149]. This work proposed a DRL policy that learned to balance the competing objectives of progress and safety in the ego-lane, comfort, and safety in the context of dynamic traffic, demonstrating that a carefully reward-shaped DRL agent can learn human-like driving behavior on a simulated highway [149].

Urban driving presents a far greater complexity than highway scenarios, requiring the agent to navigate intersections, handle occluded views, and interact with vulnerable road users. The latent deep reinforcement learning framework proposed by [150] for interpretable end-to-end urban autonomous driving is a notable contribution. The key innovation of this work is the introduction of a latent variable that captures the high-level intent of the agent, such as "turn left," "go straight," or "stop." The DRL policy learns to output both this latent intent and the low-level control commands, making the decision-making process more transparent and interpretable than a monolithic black-box policy. This latent representation also facilitates the analysis of the agent's behavior and enables a human supervisor to more easily understand and predict its actions [150].

Tactical maneuvers such as overtaking and intersection navigation have been the focus of several dedicated studies. The integration of expert guidance for learning safe overtaking, as proposed by [151], is a practical method for accelerating the learning of this challenging behavior. By first learning from a dataset of expert overtaking demonstrations via behavioral cloning and then fine-tuning with DRL, the agent learns a policy that is both sample-efficient and safe, as the expert demonstrations already encode a baseline level of caution [151]. The challenge of navigating occluded intersections, where the agent's view of cross-traffic is partially blocked by buildings or other vehicles, is tackled by [152]. Their DRL framework learns to cautiously creep forward while simultaneously estimating the probability of a hidden vehicle approaching from the occluded region, learning an effective risk-management strategy for these high-uncertainty scenarios [152]. In a complementary study, [153] proposes a more general DRL approach for autonomous vehicle intersection navigation that can handle both signalized and unsignalized intersections, learning to negotiate the right-of-way with other vehicles through a combination of discrete and continuous action outputs [153].

The validation of autonomous driving policies in high-fidelity simulation environments is a critical step before any real-world deployment. The CARLA simulator, an open-source platform for urban driving simulation, has become a standard benchmark. The study by [154] presents a comprehensive evaluation of DRL policies for autonomous driving within the CARLA environment, providing baseline results for a variety of driving scenarios and sensor configurations. This work serves as a valuable reference point for comparing the performance of different DRL algorithms on a common driving benchmark [154]. A more specialized simulation environment, F1tenth, is introduced by [155] as an open-source evaluation environment for continuous control and reinforcement learning, modeled on a small-scale autonomous racing car. While not a full driving simulation, F1tenth simulator, it provides a standardized, low-cost platform for developing and testing high-speed control policies [155].

The computational demands of training DRL policies for autonomous driving have motivated research into distributed and cloud-based architectures. The work by [156] proposes a distributed DRL system deployed on the cloud, where multiple instances of the driving simulator run in parallel on different cloud workers, each collecting experience that is asynchronously aggregated to update a shared policy. This massively parallel approach dramatically reduces the wall-clock time required to train a driving policy, enabling the learning of complex behaviors that would be infeasible on a single machine [156]. Hierarchical architectures continue to be a powerful theme in driving research. The hierarchical end-to-end autonomous driving system presented by [157] integrates Bird's Eye View (BEV) perception with DRL. The system first uses a convolutional network to generate a BEV representation of the driving scene, which shows the positions and velocities of all objects in a top-down view.

This BEV representation is then fed into a high-level DRL policy that outputs tactical commands, which are executed by a low-level continuous controller. This separation of perception and control, combined with the intermediate BEV representation, improves the interpretability and modularity of the system [157].

The challenge of learning from diverse sensor configurations is addressed by the dynamic input framework proposed by [158]. This work acknowledges that the input to a driving DRL policy, such as camera images or LiDAR point clouds, can change dynamically during a mission due to sensor failure or changing weather conditions. The framework learns a robust policy that can handle missing or corrupted input modalities, gracefully degrading its performance rather than failing catastrophically, making it more resilient for real-world deployment [158]. Finally, the use of inverse reinforcement learning (IRL) as a method for learning reward functions from human driving demonstrations, and then using these learned rewards to train an optimal policy through standard DRL, is proposed by [159] for advanced planning in autonomous vehicles. This approach allows the agent to internalize the subtle, and often unstated, preferences of human drivers, such as a preference for smooth trajectories or a tendency to bias the vehicle toward the center of the lane, resulting in more natural and acceptable driving behavior [159].

A particularly exciting sub-domain within autonomous driving is **Racing and High-Speed Control**, which pushes the physical limits of both the vehicle and the control algorithm. The work by [161] represents a landmark achievement, demonstrating champion-level drone racing using DRL. The method, which combines a model-free DRL policy with a low-level flight controller, was trained entirely in simulation using a high-fidelity physics engine and domain randomization. The learned policy was then deployed on a physical racing drone, which was able to complete a complex race course at speeds exceeding those of a world-champion human pilot, showcasing the extraordinary potential of DRL for high-speed, agile control [161]. A parallel line of research is presented in the study on autonomous drone racing by [162], which focuses on the visual navigation aspect of the task, where the drone must learn to gate detection and trajectory planning solely from camera images [162]. The problem of high-speed autonomous racing for ground vehicles is targeted by [163], which proposes a trajectory-aided DRL method. This method leverages a pre-computed, dynamically feasible racing line as a reference for the DRL policy, which then learns to deviate from this reference to optimally handle the dynamics of the car and the constraints of the track. The integration of a geometric trajectory with a learned policy provides a sample-efficient and high-performing solution for autonomous racing [163]. A critical practical consideration for racing applications is the design of the reward function, which must balance the objective of completing the course as fast as possible with the safety constraint of staying on the track. The study by [164] provides a systematic analysis of reward design and hyperparameter tuning for generalizable deep reinforcement learning agents in autonomous racing, offering practical guidelines for practitioners [164].

The domain of **Aerial Systems and Drones** extends beyond racing to encompass a wide range of autonomous capabilities. Foundational guidance for applying DRL to real-world aerial robotics is provided by the "How to train your robot" lessons learned by [165], and the Perception for Autonomous Systems (PAZ) framework by [166] provides a comprehensive software library for integrating perception modules with DRL-driven control. A critical challenge for aerial systems is the deployment of DRL on resource-constrained hardware. The FOMO (On-Device Real-Time Deep Reinforcement Learning) framework proposed by [167] directly addresses this by designing a lightweight DRL architecture that can run real-time inference on a microcontroller, enabling fully autonomous flight without any cloud connection [167].

Vision-based flight has been a major focus of research, with several works proposing methods for learning robust visuomotor policies. The monocular vision-based autonomous flight framework by [168] uses a dueling double deep Q-network to learn navigation behaviors from a single camera, demonstrating successful indoor flight [168]. A more challenging scenario is autonomous planetary landing, addressed by [169]. This work combines DRL with transfer learning, first pre-training a policy on a large corpus of simulated planetary terrain data and then fine-tuning it on a smaller set of high-fidelity landing simulations, enabling the drone to learn a policy for safe autonomous landing on unknown planetary surfaces [169]. The problem of landing on a moving platform, such as the deck of a ship, is tackled by [170] and [171]. Both works propose DRL strategies for learning the precise control actions required to track and land on a moving target, with [170] using a Gazebo-based simulation framework and [171] focusing on a more generalizable strategy [170], [171].

Research on multi-UAV systems has focused on coordination and adaptive behavior. The adaptive path planning method for multiple UAVs proposed by [173] uses a DRL framework that enables the drones to dynamically re-plan their routes in response to environmental changes or the actions of other UAVs, optimizing for collective mission objectives such as area coverage or surveillance [173]. A more efficient training methodology for single UAVs is introduced by [174], which proposes a parallel distributional prioritized DRL algorithm. By combining distributional RL, which learns the entire distribution of returns, with prioritized experience replay and parallel training, the method achieves faster convergence to a high-performing policy for UAV navigation tasks [174]. A real-world application of cooperative multi-UAV DRL is presented by [175], which focuses on autonomous decision-making for collaborative search and rescue operations. The DRL policy learns to coordinate a team of UAVs to efficiently search a designated area for a missing person or an object, dynamically allocating search areas and re-tasking drones as new information is gathered [175]. A biologically inspired approach to autonomous flight is taken by [176], which uses vulture-inspired deep reinforcement learning to reveal the principles of autonomous thermal soaring in windy conditions. The DRL policy learns to exploit rising columns of warm air (thermals) to gain altitude without expending energy, mimicking the energy-efficient soaring behavior of vultures [176].

The control of multi-rotor aerial robots is a fundamental problem addressed by [177], which proposes a DRL architecture that can learn stable and responsive motion controllers for a variety of multi-rotor configurations. A unique platform, the Autonomous Robotic Tensegrity (ART), which uses a tensegrity structure for mobility, is the subject of [178]. The DRL policy learns to control the shape of the tensegrity structure to generate locomotion, demonstrating the application of DRL to non-standard, compliant robotic platforms [178].

Marine and Underwater Systems represent a distinct set of challenges for DRL, including operation in GPS-denied environments, communication latency, and complex hydrodynamic dynamics. The problem of position tracking control for an Autonomous Underwater Vehicle (AUV) is addressed by [180], which proposes an end-to-end DRL framework using the deep deterministic policy gradient algorithm. The policy directly maps the AUV's sensor readings, such as depth, heading, and velocity, to thruster commands, enabling precise station-keeping and trajectory tracking without a dynamic model [180]. A more interactive learning paradigm is proposed by [181] for AUV path following, where a human trainer periodically provides corrective feedback or demonstrations to the DRL agent, accelerating the learning of the path-following policy.

For autonomous surface vessels, the work by [182] proposes a DRL-based tracking controller that is validated in natural waters, demonstrating the robustness of the learned policy to waves, currents, and wind. A critical safety requirement for maritime autonomous systems is compliance with the International Regulations for Preventing Collisions at Sea (COLREGs). The study by [183] directly addresses this by proposing a DRL framework that learns a COLREG-compliant collision avoidance policy, ensuring that the autonomous vessel's evasive maneuvers adhere to the established rules of the road [183].

The domain of **Human-Robot and Social Interaction** focuses on enabling robots to operate safely and socially appropriately in spaces shared with humans. The concept of shared autonomy, where a DRL policy learns to assist a human operator without fully taking control, is explored by [184]. The policy learns to infer the human's intent from their control inputs and to assist in achieving that intent, while also maintaining safety, effectively blending human and autonomous control [184]. Socially aware motion planning, which requires the robot to predict and respect the social conventions of human navigation, is the focus of [185] and [186]. The work in [185] proposes a DRL method that learns to estimate the social costs of different trajectories, such as passing too close to a pedestrian or cutting in front of a group, and plans a path that minimizes these social costs. The crowd-robot interaction method presented by [186] uses an attention-based deep reinforcement learning architecture. The robot's policy processes the positions of all nearby pedestrians through a self-attention mechanism, learning to weigh the influence of each pedestrian's future trajectory on the robot's optimal action. This enables the robot to navigate in densely crowded spaces with a high degree of social compliance [186].

The problem of learning socially appropriate robot approaching behavior towards groups is investigated by [187]. This work proposes a DRL framework that learns to approach a group of people in a socially appropriate manner, such as approaching from the front where the robot is visible, and stopping at a respectful distance. The policy learns these socially appropriate behaviors from a combination of human demonstration and trial-and-error, without being explicitly programmed with the rules of social etiquette [187]. The challenge of collision avoidance in pedestrian-rich environments is addressed by [188], which proposes a DRL method that explicitly models the interactive nature of pedestrian motion, predicting how pedestrians will react to the robot's movements. This forward-looking approach enables the robot to anticipate and avoid collisions more effectively than a purely reactive policy [188].

Collaborative human-robot interaction is further explored by [189], which presents a framework for human-centered collaborative robots that use DRL. The framework addresses the problem of a robot learning to work alongside a human on a shared task, such as lifting and moving a heavy object, where the robot must learn to adapt its behavior to the human's actions in real-time. The cycle-of-learning framework proposed by [190] directly addresses this by using DRL to learn from human demonstrations and corrective feedback in a continuous loop. The robot learns from the human, then attempts the task, and then receives corrective feedback from the human to refine its policy, enabling efficient and safe learning of collaborative behaviors [190]. The efficient combination of human demonstrations and interventions for safe training has been previously discussed as a safety mechanism [132], and this same theme is echoed in the context of autonomous driving by [151]. A specific industrial application of human-robot collaboration is presented by [192], which proposes a prediction-based path planning method for safe and efficient human-robot collaboration in construction tasks. The DRL framework predicts the future movements of human workers based on their current motion and planned tasks, and then plans the robot's path to avoid interfering with those movements, improving both safety and productivity [192].

Legged, Bio-inspired, and Micro-Aerial robotics represents a frontier where DRL is used to master the complex dynamics of multi-limbed locomotion and unconventional robot morphologies. The work on robust biped locomotion by [219] proposes a powerful synergistic approach, where a model-based analytical controller provides low-level stability and basic gait generation, while a DRL policy learns to modulate the parameters of this controller to achieve robust walking over uneven terrain and recover from pushes. This hierarchical decomposition combines the formal guarantees of analytical control with the adaptability of DRL [219]. The feedback control system for the Cassie biped robot, as presented by [194], is a similar endeavor, using DRL to learn a full-body feedback policy that can maintain balance and walk dynamically, representing a state-of-the-art application of DRL to a complex humanoid platform [194]. The problem of achieving human-like motion on a quadrupedal robot is the subject of [195], which proposes the Humanconquad framework. A human operator's joint motion is captured and retargeted to the quadruped, and a DRL policy is trained to mimic this motion while maintaining balance in dynamic situations, effectively "humanizing" the quadruped's gait [195].

The challenge of designing and validating DRL systems for modular legged robots is addressed by [196], which presents an automated DRL environment for a modular legged robot. This environmental framework allows for rapid prototyping and testing of DRL policies for different leg configurations, accelerating research in this area. An alternative, biologically-plausible paradigm for legged locomotion is explored by [197], which proposes learning to walk using spike-based reinforcement learning for central pattern generation (CPG) on a hexapod robot. Instead of a standard DNN, the policy is implemented using a spiking neural network, which is considered to be more energy-efficient and biologically realistic. The spike-timing-dependent plasticity (STDP) learning rule is used to learn the weights of the CPG network, demonstrating that a principled, neuromorphic approach can achieve effective locomotion [197].

Medical, Surgical, and Inspection applications of DRL are among the most impactful and safety-critical. The semi-autonomous control framework for robotic surgery proposed by [198] uses DRL to learn an assistive policy that can execute low-level surgical tasks, such as knot-tying or suturing, while the surgeon retains high-level control over the procedure. The DRL policy helps to stabilize the surgeon's motions and perform precise, repetitive movements, reducing surgeon fatigue and improving outcome consistency [198]. The specific problem of autonomous navigation of an ultrasound probe is addressed by [199] and [200]. The work in [199] proposes a DRL framework that learns to navigate the ultrasound probe to acquire standard diagnostic scan planes, such as the four-chamber view of the heart. The policy learns the visuomotor associations between the image seen on the ultrasound machine and the required probe manipulation [199]. The ultrasound-guided robotic navigation system in [200] extends this with a more general DRL policy for probe navigation [200]. In a different medical context, the use of machine learning for robotic and autonomous inspection of mechanical systems and civil infrastructure is explored by [201], which surveys the broader applicability of DRL to inspection tasks, such as inspecting pipelines, bridges, or aircraft components [201].

Industrial and Construction applications leverage DRL for tasks that require high precision, adaptability, and autonomy. The autonomous robotic additive manufacturing framework proposed by [202] uses distributed model-free deep reinforcement learning to control the deposition of material in a 3D printing process. The DRL policy learns to manage the complex material flow and structural stability constraints in real time, enabling the robot to build complex, custom structures without human intervention [202]. The problem of real-time assembly planning in robot-based prefabricated construction is addressed by [203]. This work proposes a DRL method that learns to sequence the assembly of prefabricated building components, such as wall panels or structural beams, in an optimal order, taking into account the robot's reach, the component's geometry, and the stability of the partially assembled structure [203]. A more traditional industrial application is presented by [204], which proposes a wheel loader scooping controller using DRL. The policy learns to optimally control the bucket trajectory to efficiently scoop up a pile of gravel or soil, a task that requires precise control of hydraulic forces and estimation of the load's weight [204].

The application of DRL to **Autonomous Defense and Security** represents a niche but important frontier. The study by [205] provides a strategic overview of the role of reinforcement learning in autonomous defense systems, discussing potential applications such as autonomous drone swarms for surveillance and interception, and the unique ethical and safety challenges these deployment challenges [205].

Finally, the category of **Specialized Emerging Frontiers** captures the most innovative and unconventional research directions within our corpus. The integration of quantum computing with DRL is a bold new direction. The work by [206] proposes a quantum deep reinforcement learning algorithm for robot navigation tasks. Instead of using classical deep neural networks, the policy is implemented using a parameterized quantum circuit (a quantum neural network). The quantum DRL agent is shown to learn navigation policies with a significantly smaller number of parameters compared to a classical DRL agent, suggesting that quantum computing could offer a more efficient representation for DRL policies [206]. A further evolution of this concept is the quantum-enhanced hybrid reinforcement learning framework for dynamic path planning proposed by [207]. This framework combines a classical DRL agent with a quantum processing unit that is used to solve a subset of the optimization problem encountered during planning, accelerating the decision-making process [207]. The exploration of spiking neural networks (SNNs) for DRL in robotic tasks, as pursued by [208], represents a fundamentally different computational paradigm. SNNs are event-driven and biologically plausible, offering the potential for ultra-low-power operation on neuromorphic hardware. The study demonstrates that an SNN-based DRL agent can learn to perform control tasks, such as balancing a pole, with performance comparable to a standard DNN-based agent, opening the door to energy-efficient DRL for edge robots [208].

The deployment of DRL on resource-constrained edge devices is also the impetus behind the robust iterative value conversion framework for neurochip-driven edge robots presented by [209]. This work proposes a technique for converting a pre-trained DRL policy into a form that can be executed on a custom neurochip, which is a specialized hardware accelerator for neural networks. The conversion process includes a robust iterative optimization stage that ensures the DRL policy's performance is preserved after the conversion, despite the reduced numerical precision of the neurochip [209]. The application of DRL to the autonomous Internet of Things (IoT) is explored by [210], which examines the model, applications, and challenges of using DRL to control a network of IoT devices, such as smart sensors and actuators, for tasks like resource allocation and traffic optimization [210]. Finally, the modular task allocation framework for autonomous mobile systems by [211] proposes a modular task allocation framework that uses DRL to dynamically decompose a high-level mission into a sequence of subtasks and to allocate these subtasks to the available autonomous agents in the network [211].

Several works in this dimension do not fit neatly into a single application domain but instead contribute to the foundational or theoretical understanding of DRL in autonomous systems. These include the review of neurocomputing applications in autonomous systems by [212], the discussion of stochastic models for autonomous systems by [213], and the funneled Bayesian optimization framework for the design, tuning, and control of autonomous systems by [214]. The study on the societal implications of deep reinforcement learning by [217] provides a crucial, reflective analysis of the broader consequences of this technology. The adaptive hybrid algorithms for real-time decision-making in autonomous systems, presented in [191], propose a general purpose framework that combines DRL with other AI techniques. The comprehensive review of reinforcement learning applications in autonomous systems, spanning from traffic optimization to robotics, by [215], and the roadmap for the theory to deployment of deep learning and reinforcement learning for autonomous unmanned aerial systems by [216], both serve as valuable overviews that contextualize the more specialized works in the field. Finally, the reinforcement learning behavioral control method for nonlinear autonomous systems by [218] proposes a general control architecture for any autonomous system with nonlinear dynamics.

Discussion

The synthesis of 210 studies across seven research dimensions reveals a landscape of deep reinforcement learning for autonomous systems and robotics that is simultaneously vibrant, fragmented, and marked by persistent gaps between methodological ambition and practical deployment. Taken together, the findings presented in the results section coalesce around several overarching themes that transcend individual sub-domains. The first and perhaps most prominent pattern that emerges across the literature is the clear trajectory from foundational algorithmic development toward hybrid and integrated methodologies. The early works in our corpus, concentrated around 2016 and 2017, were primarily concerned with establishing the viability of DRL as a paradigm for robotics, demonstrating that methods like DQN and DDPG could learn sensorimotor policies from scratch in simulated environments. However, as the field matured, a consistent shift toward hybridization becomes apparent. From 2019 onward, a growing number of studies have moved beyond monolithic end-to-end DRL and have instead embraced frameworks that combine model-free DRL with model-based planning [87], [145], hierarchical decomposition [134], [220], and classical control theory [126], [193]. This hybridization is not merely an academic exercise; it consistently yields tangible benefits in terms of sample efficiency, safety, and generalization, which are precisely the bottlenecks that plague pure model-free DRL in real-world applications.

A second critical insight that emerges across the reviewed studies is the stark disparity in research effort between performance-oriented advancements and safety-oriented validation. The sheer volume of work in the Autonomous Navigation and Path Planning dimension, comprising 55 studies, stands in stark contrast to the Safety, Verification, and Robustness dimension, which contains only 19 studies. This imbalance is concerning, particularly given that many of the navigation and safety-related studies we did identify explicitly demonstrate the vulnerability of standard DRL policies to adversarial perturbations [128], [129] or distributional shift [120]. We consistently found that while the performance of DRL agents on standard benchmarks continues to improve, the rigorous verification of these same policies remains a predominantly open challenge. The formal verification methods that do exist, such as reachability analysis or SMT-based verification of neural networks [116], [118], are typically computationally intensive and do not scale to the high-dimensional state spaces or deep architectures that are standard in modern DRL. This gap between performance and assurance represents a fundamental roadblock to the deployment of DRL in safety-critical autonomous systems, such as autonomous driving or surgical robotics, where a certified safety guarantee, rather than a high average success rate, is a prerequisite for regulatory approval.

The multi-agent and swarm dimension of the literature reveals a third overarching pattern: a tension between the scalability of fully decentralized approaches and the stability of centralized or communication-based coordination. The studies on decentralized collision avoidance [97], [115] demonstrate impressive emergent coordination without any explicit communication, which is highly desirable for scalability. However, these same works consistently report challenges with convergence and performance degradation as the number of agents increases, or as the environment becomes more crowded and dynamic. In contrast, the works that incorporate a centralized training, decentralized execution (CTDE) framework [104], [113] exhibit more stable learning and higher final performance, but they introduce a communication infrastructure and a central point of failure that may not be feasible in large-scale or adversarial settings. This trade-off between scalability and coordination quality is a recurring theme across the 23 multi-agent studies, and no single strategy appears to resolve it universally. The choice between fully decentralized and CTDE architectures must therefore be made carefully, contingent on the specific application's tolerance for communication overhead, the criticality of optimal coordination, and the expected scale of the robot team.

Turning to the implications of the findings are substantial and cut across theoretical andpractical dimensions. From a theoretical perspective, the consistent success of hybrid and hierarchical frameworks challenges the prevailing assumption that end-to-end model-free DRL is the most general or powerful paradigm for robotic control. The evidence from our review suggests that incorporating structure—whether through a model of dynamics, a hierarchy of subpolicies, or a set of safety constraints derived from physics—not only improves sample efficiency but also enhances the interpretability and transferability of learned policies. This finding has implications for how we conceptualize learning in autonomous systems. It implies that the optimal path to robust autonomy may not be through ever-larger neural networks trained with ever-more data, but rather through principled integrations of learning with existing mechanistic knowledge. The theoretical contribution of our synthesis is therefore to reframe the research question from "how can we make DRL work?" to "how should we compose and constrain DRL with other computational structures to make it work reliably?"

The practical implications of our review are equally compelling. For practitioners and engineers deploying DRL in real-world systems, our findings strongly caution against the naive application of off-the-shelf DRL algorithms directly to physical hardware. The literature consistently demonstrates that successful real-world deployment requires a holistic engineering approach that goes far beyond the algorithmic core. The studies. This includes careful simulation and domain randomization [29], [59], robust systems integration through frameworks like Arena-Rosnav or Pic4rl-gym [58], [61], and the incorporation of human oversight or safety interpolation mechanisms [126], [133]. Furthermore, the domain-specific applications in autonomous driving and medical robotics underscore that the reward function design, rather than the algorithmic architecture, is often the single most impactful factor in determining the quality and safety of the learned behavior. Policymakers and regulators developing standards for autonomous systems should take note of the significant gap between algorithmic performance and formal verification; our review suggests that current DRL methods are not yet ready for safety-critical deployment without substantial, manually engineered safety overlays.

We must acknowledge the methodological limitations of this systematic review, which may influence the conclusions we have drawn. The primary limitation stems from the scope and design of our search strategy. Despite our use of five major databases and systematic search strings, we may have inadvertently missed relevant studies that use terminology outside our search parameters, such as those that focus on specific control-theoretic applications without explicitly mentioning "deep reinforcement learning" in the title or abstract. This selection bias could skew our findings toward the more established and explicitly named methods, potentially underrepresenting niche or emerging applications that have not yet adopted standard DRL terminology. Additionally, the exclusion of non-English language studies introduces a language bias, which may particularly affect the representation of research from non-Anglophone robotics communities.

Another limitation is the inherent subjectivity in our classification scheme. While we developed seven research dimensions iteratively through pilot reading, the assignment of individual studies to a single dimension is often an oversimplification. Many papers, particularly those on advanced learning paradigms or safety, contribute to multiple dimensions simultaneously. This categorical rigidity may obscure some of the cross-cutting relationships between, for example, a hierarchical DRL method for manipulation that also incorporates safety constraints. The qualitative synthesis approach, while providing rich insight, also precludes any formal statistical meta-analysis. We cannot quantitatively measure the relative effect sizes of different algorithmic innovations across studies, as the reported metrics, environments, and tasks are too heterogeneous to allow for direct comparison.

Finally, there is an inherent publication bias in any systematic review. Studies that report positive results, such as a new algorithm outperforming a baseline, are more likely to be published and indexed than those reporting negative or null results. This can create an overly optimistic picture of the field's progress. The decline in publication volume from 2022 onward, which we noted in our research trends, could be partially attributable to a "crowding out" effect, where successful methods are more likely to be published, while incremental or negative findings are increasingly filed away. These limitations do not invalidate our findings, but they contextualize them, suggesting that the field, while advanced, may be less uniformly successful and more fraught with unreported failures than the published record suggests.

Based on the gaps, inconsistencies, and limitations uncovered by our review, several promising directions for future research emerge. There is a clear and urgent need for research that explicitly integrates formal verification methods with scalable DRL training processes. Future research should explore methods for training policies that are inherently verifiable, for instance by constraining the network architecture to be more amenable to reachability analysis, or by incorporating proof-producing certificates into the learning objective itself. The safety verification problem is not merely an afterthought; it is the central scientific challenge that DRL must overcome to transition from a simulation curiosity to an engineering standard. The development of efficient, scalable verification tools for neural network policies should be prioritized as a first-class research goal, perhaps leveraging advances in abstract interpretation or differentiable logics.

Another critical area for future investigation is the improvement of sample efficiency for real-world systems. While model-based DRL and hierarchical RL offer promising avenues, the gap between the millions of samples required by DRL and the thousand or so interactions that are practical on physical hardware remains immense. understudied areas include the use of strong inductive biases from physics or geometry, as seen in the physics-regulated invariant embeddings study [124], and the deep integration of prior knowledge through Bayesian controller fusion [144]. We suggest that future research should focus on learning from a handful of demonstrations, and then rapidly fine-tuning with a limited number of safe real-world interactions, rather than the current paradigm of training from scratch in simulation and hoping for zero-shot transfer.

Finally, the domain of multi-agent DRL would benefit greatly from a deeper theoretical understanding of convergence in decentralized settings. The empirical successes of decentralized collision avoidance are encouraging, but the lack of theoretical guarantees on convergence or the quality of the emergent equilibrium is a significant weakness. Future research should explore game-theoretic analyses of decentralized DRL, perhaps drawing on methods from mean-field game theory to analyze swarm-scale interactions, or developing mechanisms for agent-to-agent communication that are learned, efficient, and robust to adversarial corruption.

Conclusion

This systematic literature review synthesized 210 studies to provide a comprehensive mapping of deep reinforcement learning techniques for autonomous systems and robotics, addressing the gap between algorithmic proliferation and holistic understanding. We found that the field has transitioned from foundational algorithm development toward hybrid methodologies that integrate model-based planning, hierarchical control, and classical safety mechanisms. However, our analysis revealed a persistent and concerning asymmetry: performance-oriented research on navigation and manipulation far outweighs work on safety verification and robustness, creating a critical bottleneck for real-world deployment. Furthermore, distributed learning in multi-agent settings continues to struggle with fundamental convergence challenges, and sample efficiency remains insufficient for routine physical hardware training.

The theoretical implication of our synthesis is that future progress will depend not on scaling monolithic DRL architectures, but on principled compositions of learning compositions that embed domain knowledge and formal constraints into the training process itself. For practitioners, our findings underscore that successful deployment requires comprehensive systems engineering beyond algorithmic choice—encompassing reward design, simulation fidelity, and safety overlays. Looking forward, we identify three urgent research priorities: scalable formal verification methods integrated with learning, sample-efficient paradigms leveraging strong inductive biases, and theoretical convergence guarantees for decentralized multi-agent coordination. Closing these gaps will be essential for transforming DRL from a simulation-oriented tool into a reliable foundation for real-world autonomous systems.

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, *et al.*, "Human-level control through deep reinforcement learning," *nature*, 2015.

[2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, *et al.*, "Mastering the game of go without human knowledge," *nature*, 2017.

[3] O. Andrychowicz, B. Baker, *et al.*, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, 2020.

[4] Y. Zhu, R. Mottaghi, E. Kolve, J. Lim, *et al.*, "Target-driven visual navigation in indoor scenes using deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2017.

[5] R. Sutton and A. Barto, "Reinforcement learning: An introduction," cambridge.org, 1998.

[6] M. Bojarski, D. D. Testa, D. Dworakowski, *et al.*, "End to end learning for self-driving cars," arXiv preprint arXiv:1604.07316, 2016.

[7] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, 2016.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, *et al.*, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[9] V. Mnih, A. Badia, M. Mirza, A. Graves, *et al.*, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*, 2016.

[10] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *IEEE/RSJ international conference on intelligent robots and systems*, 2012.

[11] X. Wang, S. Wang, X. Liang, D. Zhao, *et al.*, "Deep reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[12] J. Achiam, D. Held, A. Tamar, *et al.*, "Constrained policy optimization," in *International conference on machine learning*, 2017.

[13] L. Kaiser, M. Babaeizadeh, P. Milos, B. Osinski, *et al.*, "Model-based reinforcement learning for atari," arXiv preprint arXiv:1903.00374, 2019.

[14] D. Han, B. Mulyana, V. Stankovic, and S. Cheng, "A survey on deep reinforcement learning algorithms for robotic manipulation," *Sensors*, 2023.

[15] B. Kiran, I. Sobh, V. Talpaert, P. Mannion, *et al.*, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[16] M. Page, J. McKenzie, P. Bossuyt, *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[17] S. Mousavi, M. Schukat, and E. Howley, "Deep reinforcement learning: An overview," in *Proceedings of SAI intelligent systems conference*, 2016.

[18] Y. Fenjiro and H. Benbrahim, "Deep reinforcement learning overview of the state of the art," *Journal of Automation, Mobile Robotics and Intelligent Systems*, 2018.

[19] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, 2019.

[20] F. AlMahamid and K. Grolinger, "Reinforcement learning algorithms: An overview and classification," in *2021 IEEE canadian conference on electrical and computer engineering*, 2021.

[21] M. Hillebrand, M. Lakhani, and R. Dumitrescu, "A design methodology for deep reinforcement learning in autonomous systems," *Procedia Manufacturing*, 2020.

[22] Y. Wang, M. Hou, K. Plataniotis, *et al.*, "Towards a theoretical framework of autonomous systems underpinned by intelligence and systems sciences," *IEEE/CAA Journal of Automatica Sinica*, 2020.

[23] H. Zhu, J. Yu, A. Gupta, D. Shah, K. Hartikainen, *et al.*, "The ingredients of real-world robotic reinforcement learning," arXiv preprint arXiv:2004.12570, 2020.

[24] A. Rayhan, "Artificial intelligence in robotics: From automation to autonomous systems," *IEEE Transactions on Robotics*, 2023.

[25] A. Azar and A. Koubaa, "Artificial intelligence for robotics and autonomous systems applications," *Springer*, 2023.

[26] C. Thompson, R. Talla, J. Gummadi, *et al.*, "Reinforcement learning techniques for autonomous robotics," *Asian Journal of Research in Computer Science*, 2019.

[27] M. Ejaz, T. Tang, and C. Lu, "Autonomous visual navigation using deep reinforcement learning: An overview," in *2019 IEEE student conference on research and development*, 2019.

[28] E. Marchesini and A. Farinelli, "Genetic deep reinforcement learning for mapless navigation," in *Proceedings of the 19th international conference on autonomous agents and multiagent systems*, 2020.

[29] L. Tai, G. Paolo, and M. Liu, "Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation," … on intelligent robots; systems, 2017.

[30] M. Dobrevski and D. Skočaj, "Deep reinforcement learning for map-less goal-driven robot navigation," *International Journal of Advanced Robotic Systems*, 2021.

[31] O. Zhelo, J. Zhang, L. Tai, M. Liu, and W. Burgard, "Curiosity-driven exploration for mapless navigation with deep reinforcement learning," arXiv preprint arXiv:1804.00456, 2018.

[32] M. Pfeiffer, S. Shukla, M. Turchetta, *et al.*, "Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations," *IEEE Robotics and Automation Letters*, 2018.

[33] R. Grando, J. de Jesus, V. Kich, *et al.*, "Deep reinforcement learning for mapless navigation of a hybrid aerial underwater vehicle with medium transition," in *IEEE international conference on robotics and automation*, 2021.

[34] A. Singla, S. Padakandla, *et al.*, "Memory-based deep reinforcement learning for obstacle avoidance in UAV with limited environment knowledge," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[35] Z. Zhang, H. Fu, J. Yang, and Y. Lin, "Deep reinforcement learning for path planning of autonomous mobile robots in complicated environments," *Complex & Intelligent Systems*, 2025.

[36] H. Liu, Y. Shen, S. Yu, Z. Gao, and T. Wu, "Deep reinforcement learning for mobile robot path planning," arXiv preprint arXiv:2404.06974, 2024.

[37] L. Yang, J. Bi, and H. Yuan, "Dynamic path planning for mobile robots with deep reinforcement learning," *IFAC-PapersOnLine*, 2022.

[38] A. Kumaar and S. Kochuvila, "Mobile service robot path planning using deep reinforcement learning," *IEEE Access*, 2023.

[39] G. Chen, L. Pan, Y. Chen, P. Xu, Z. Wang, *et al.*, "Robot navigation with map-based deep reinforcement learning," in *IEEE international conference on sensing and control*, 2020.

[40] V. Miranda, A. Neto, G. Freitas, *et al.*, "Generalization in deep reinforcement learning for robotic navigation by reward shaping," *IEEE Transactions on Cognitive and Developmental Systems*, 2023.

[41] K. Honda, R. Yonetani, M. Nishimura, *et al.*, "When to replan? An adaptive replanning strategy for autonomous navigation using deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2024.

[42] J. Hu, H. Niu, J. Carrasco, B. Lennox, *et al.*, "Voronoi-based multi-robot autonomous exploration in unknown environments via deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, 2020.

[43] J. Zhang, J. Springenberg, J. Boedecker, *et al.*, "Deep reinforcement learning with successor features for navigation across similar environments," in *IEEE international conference on robotics and automation*, 2017.

[44] A. bin Kamarulariffin, A. Ibrahim, *et al.*, "Improving deep reinforcement learning training convergence using fuzzy logic for autonomous mobile robot navigation." *International Journal of Fuzzy Systems*, 2023.

[45] Y. Xu, J. Yu, J. Tang, J. Qiu, J. Wang, *et al.*, "Explore-bench: Data sets, metrics and evaluations for frontier-based and deep-reinforcement-learning-based autonomous exploration," in *IEEE international conference on robotics and automation*, 2022.

[46] R. Cimurs, I. Suh, and J. Lee, "Goal-driven autonomous exploration through deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2021.

[47] L. Tai and M. Liu, "Towards cognitive exploration through deep reinforcement learning for mobile robots," arXiv preprint arXiv:1610.01733, 2016.

[48] F. Chen, J. Martin, Y. Huang, J. Wang, *et al.*, "Autonomous exploration under uncertainty via deep reinforcement learning on graphs," in *IEEE/RSJ international conference on intelligent robots and systems*, 2020.

[49] Y. Cao, R. Zhao, Y. Wang, B. Xiang, *et al.*, "Deep reinforcement learning-based large-scale robot exploration," *IEEE Robotics and Automation Letters*, 2024.

[50] F. Chen, S. Bai, T. Shan, and B. Englot, "Self-learning exploration and mapping for mobile robots via deep reinforcement learning," *Aiaa scitech 2019 forum*, 2019.

[51] M. Sarfi and M. Bisheban, "Risk-sensitive autonomous exploration of unknown environments: A deep reinforcement learning perspective," *Journal of Intelligent & Robotic Systems*, 2025.

[52] J. Li, M. Ran, H. Wang, and L. Xie, "A behavior-based mobile robot navigation method with deep reinforcement learning," *Unmanned Systems*, 2021.

[53] P. Wenzel, T. Schön, L. Leal-Taixé, *et al.*, "Vision-based mobile robotics obstacle avoidance with deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2021.

[54] G. Xu, W. Jiang, Z. Wang, and Y. Wang, "Autonomous obstacle avoidance and target tracking of UAV based on deep reinforcement learning," *Journal of Intelligent & Robotic Systems*, 2022.

[55] L. Kästner, C. Marx, and J. Lambrecht, "Deep-reinforcement-learning-based semantic navigation of mobile robots in dynamic environments," in *2020 IEEE 16th international conference on automation science and engineering*, 2020.

[56] S. Josef and A. Degani, "Deep reinforcement learning for safe local planning of a ground vehicle in unknown rough terrain," *IEEE Robotics and Automation Letters*, 2020.

[57] K. Weerakoon, A. Sathyamoorthy, *et al.*, "Terp: Reliable planning in uneven outdoor environments using deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2022.

[58] L. Kästner, T. Buiyan, L. Jiao, T. Le, *et al.*, "Arena-rosnav: Towards deployment of deep-reinforcement-learning-based obstacle avoidance into conventional autonomous navigation systems," in *IEEE/RSJ international conference on intelligent robots and systems*, 2021.

[59] H. Hu, K. Zhang, A. Tan, M. Ruan, *et al.*, "A sim-to-real pipeline for deep reinforcement learning for autonomous robot navigation in cluttered rough terrain," *IEEE Robotics and Automation Letters*, 2021.

[60] S. Krishnan, B. Boroujerdian, W. Fu, A. Faust, and V. Reddi, "Air learning: A deep reinforcement learning gym for autonomous aerial robot visual navigation," *Machine Learning*, 2021.

[61] M. Martini, A. Eirale, S. Cerrato, *et al.*, "Pic4rl-gym: A ros2 modular framework for robots autonomous navigation with deep reinforcement learning," in *2023 3rd international conference on autonomous systems and robotics*, 2023.

[62] M. Luong and C. Pham, "Incremental learning for autonomous navigation of mobile robots based on deep reinforcement learning," *Journal of Intelligent & Robotic Systems*, 2021.

[63] G. Kahn, A. Villaflor, B. Ding, P. Abbeel, *et al.*, "Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation," in *IEEE international conference on robotics and automation*, 2018.

[64] H. Surmann, C. Jestel, R. Marchel, F. Musberg, *et al.*, "Deep reinforcement learning for real autonomous mobile robot navigation in indoor environments," arXiv preprint arXiv:2005.13857, 2020.

[65] Y. Xi, "Research on autonomous mobile robot navigation technology based on deep reinforcement learning," *Highlights Sci. Eng. Technol*, 2024.

[66] H. Quan, Y. Li, and Y. Zhang, "A novel mobile robot navigation method based on deep reinforcement learning," *International Journal of Advanced Robotic Systems*, 2020.

[67] T. Manh, C. Manh, D. Tien, M. Van, *et al.*, "Autonomous navigation for omnidirectional robot based on deep reinforcement learning," *Journal of Mechanical Science and Technology*, 2020.

[68] M. Martini, S. Cerrato, F. Salvetti, *et al.*, "Position-agnostic autonomous navigation in vineyards with deep reinforcement learning," in *2022 IEEE 18th international conference on automation science and engineering*, 2022.

[69] L. Xu, H. Liu, H. Zhao, T. Zheng, T. Jiang, *et al.*, "Autonomous navigation of unmanned vehicle through deep reinforcement learning," in *Proceedings of the 5th international conference on autonomous agents and multiagent systems*, 2024.

[70] O. Doukhi and D. Lee, "Deep reinforcement learning for autonomous map-less navigation of a flying robot," *Ieee Access*, 2022.

[71] V. Hodge, R. Hawkins, and R. Alexander, "Deep reinforcement learning for drone navigation using sensor data," *Neural Computing and Applications*, 2021.

[72] R. Polvara, M. Patacchiola, S. Sharma, J. Wan, *et al.*, "Autonomous quadrotor landing using deep reinforcement learning," arXiv preprint arXiv:1709.03339, 2017.

[73] C. Yan, X. Xiang, and C. Wang, "Towards real-time path planning through deep reinforcement learning for a UAV in dynamic environments," *Journal of Intelligent & Robotic Systems*, 2020.

[74] G. Kim, "Cooperative deep reinforcement learning policies for autonomous navigation in complex environments," *IEEE Access*, 2024.

[75] W. Li, M. Yue, J. Shangguan, and Y. Jin, "Navigation of mobile robots based on deep reinforcement learning: Reward function optimization and knowledge transfer," *International Journal of Control, Automation and Systems*, 2023.

[76] X. Wang, Y. Sun, Y. Xie, J. Bin, and J. Xiao, "Deep reinforcement learning-aided autonomous navigation with landmark generators," *Frontiers in Neurorobotics*, 2023.

[77] Z. Xu, B. Liu, X. Xiao, A. Nair, and P. Stone, "Benchmarking reinforcement learning techniques for autonomous navigation," arXiv preprint arXiv:2210.04839, 2022.

[78] S. Amarjyoti, "Deep reinforcement learning for robotic manipulation-the state of the art," arXiv preprint arXiv:1701.08878, 2017.

[79] S. Gu, E. Holly, T. Lillicrap, *et al.*, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *IEEE international conference on robotics and automation*, 2017.

[80] H. Zhu, A. Gupta, A. Rajeswaran, *et al.*, "Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost," in *IEEE international conference on robotics and automation*, 2019.

[81] S. Joshi, S. Kumra, and F. Sahin, "Robotic grasping using deep reinforcement learning," in *2020 IEEE 16th international conference on automation science and engineering*, 2020.

[82] M. Breyer, F. Furrer, T. Novkovic, *et al.*, "Comparing task simplifications to learn closed-loop object picking using deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2019.

[83] Y. Chen, X. Li, S. Guo, X. Ng, and M. Ang, "Real2sim or sim2real: Robotics visual insertion using deep reinforcement learning and real2sim policy adaptation," in *International conference on simulation, modeling, and programming for autonomous robots*, 2022.

[84] A. Orsula, S. Bøgh, M. Olivares-Mendez, *et al.*, "Learning to grasp on the moon from 3d octree observations with deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2022.

[85] W. Hu, B. Huang, W. Lee, S. Yang, Y. Zheng, *et al.*, "Dexterous in-hand manipulation of slender cylindrical objects through deep reinforcement learning with tactile sensing," *Robotics and Autonomous Systems*, 2025.

[86] F. Zhang, J. Leitner, M. Milford, B. Upcroft, *et al.*, "Towards vision-based deep reinforcement learning for robotic motion control," arXiv preprint arXiv:1511.03791, 2015.

[87] F. Ebert, C. Finn, S. Dasari, A. Xie, A. Lee, *et al.*, "Visual foresight: Model-based deep reinforcement learning for vision-based robotic control," arXiv preprint arXiv:1812.00568, 2018.

[88] G. Schoettler, A. Nair, J. Luo, S. Bahl, *et al.*, "Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards," in *IEEE/RSJ international conference on intelligent robots and systems*, 2020.

[89] A. Franceschetti, E. Tosello, N. Castaman, *et al.*, "Robotic arm control and task training through deep reinforcement learning," in *International conference on intelligent robotics and applications*, 2021.

[90] D. Honerkamp, T. Welschehold, *et al.*, "Learning kinematic feasibility for mobile manipulation through deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2021.

[91] G. Kang, H. Seong, D. Lee, and D. Shim, "A versatile door opening system with mobile manipulator through adaptive position-force control and reinforcement learning," *Robotics and Autonomous Systems*, 2024.

[92] N. Nguyen, T. Nguyen, S. Nahavandi, A. Bhatti, *et al.*, “Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery,” arXiv preprint arXiv:1902.05183, 2019.

[93] T. Ho, K. Nguyen, and M. Cheriet, “Federated deep reinforcement learning for task scheduling in heterogeneous autonomous robotic system,” *IEEE Transactions on Automation Science and Engineering*, 2022.

[94] L. Zhu, J. Cheng, H. Zhang, Z. Cui, *et al.*, “Autonomous and adaptive role selection for multi-robot collaborative area search based on deep reinforcement learning,” *IEEE Transactions on Automation Science and Engineering*, 2025.

[95] K. Shibata, T. Jimbo, and T. Matsubara, “Deep reinforcement learning of event-triggered communication and consensus-based control for distributed cooperative transport,” *Robotics and Autonomous Systems*, 2023.

[96] M. Everett, Y. Chen, and J. How, “Motion planning among dynamic, decision-making agents with deep reinforcement learning,” in *IEEE/RSJ international conference on intelligent robots and systems*, 2018.

[97] T. Fan, P. Long, W. Liu, and J. Pan, “Fully distributed multi-robot collision avoidance via deep reinforcement learning for safe and efficient navigation in complex scenarios,” arXiv preprint arXiv:1808.03841, 2018.

[98] T. Fan, P. Long, W. Liu, and J. Pan, “Distributed multi-robot collision avoidance via deep reinforcement learning for navigation in complex scenarios,” *International Journal of Robotics Research*, 2020.

[99] Y. Chen, M. Liu, M. Everett, *et al.*, “Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning,” in *IEEE international conference on systems, man, and cybernetics*, 2017.

[100] P. Long, T. Fan, X. Liao, W. Liu, *et al.*, “Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning,” in *IEEE international conference on robotics and automation*, 2018.

[101] R. Han, S. Chen, and Q. Hao, “Cooperative multi-robot navigation in dynamic environment with deep reinforcement learning,” in *2020 IEEE international conference on robotics and automation*, 2020.

[102] S. Liu, P. Chang, W. Liang, N. Chakraborty, *et al.*, “Decentralized structural-rnn for robot crowd navigation with deep reinforcement learning,” in *IEEE international conference on robotics and automation*, 2021.

[103] Y. Yang, L. Juntao, and P. Lingling, “Multi-robot path planning based on a deep reinforcement learning DQN algorithm,” *CAAI Transactions on Intelligence Technology*, 2020.

[104] A. Viseras and R. Garcia, “DeepIG: Multi-robot information gathering with deep reinforcement learning,” *IEEE Robotics and Automation Letters*, 2019.

[105] A. Tan, F. Bejarano, Y. Zhu, R. Ren, *et al.*, "Deep reinforcement learning for decentralized multi-robot exploration with macro actions," *IEEE Robotics and Automation Letters*, 2022.

[106] M. Hüttenrauch, A. Šošić, and G. Neumann, "Guided deep reinforcement learning for swarm systems," arXiv preprint arXiv:1709.06011, 2017.

[107] S. Na, H. Niu, B. Lennox, and F. Arvin, "Bio-inspired collision avoidance in swarm systems via deep reinforcement learning," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.

[108] F. Martínez, H. Montiel, and L. Wanumen, "A deep reinforcement learning strategy for autonomous robot flocking." *International Journal of Advanced Robotic Systems*, 2023.

[109] L. Zhang, Y. Sun, A. Barth, and O. Ma, "Decentralized control of multi-robot system in cooperative object transportation using deep reinforcement learning," *IEEE access*, 2020.

[110] C. D. Souza, R. Newbury, A. Cosgun, *et al.*, "Decentralized multi-agent pursuit using deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2021.

[111] J. Ma, H. Lu, J. Xiao, Z. Zeng, and Z. Zheng, "Multi-robot target encirclement control with collision avoidance via deep reinforcement learning," *Journal of Intelligent & Robotic Systems*, 2020.

[112] J. Hook, S. El-Sedky, V. D. Silva, and A. Kondoz, "Learning data-driven decision-making policies in multi-agent environments for autonomous systems," *Cognitive Systems Research*, 2021.

[113] J. Gupta, M. Egorov, and M. Kochenderfer, "Cooperative multi-agent control using deep reinforcement learning," in *International conference on neural information processing*, 2017.

[114] N. Nguyen, T. Nguyen, and S. Nahavandi, "Multi-agent behavioral control system using deep reinforcement learning," *Neurocomputing*, 2019.

[115] J. Lin, X. Yang, P. Zheng, and H. Cheng, "End-to-end decentralized multi-robot navigation in unknown complex environments via deep reinforcement learning," in *2019 IEEE international conference on robotics and automation*, 2019.

[116] Y. Dong, X. Zhao, S. Wang, *et al.*, "Reachability verification based reliability assessment for deep reinforcement learning controlled robotics and autonomous systems," *IEEE Robotics and Automation Letters*, 2024.

[117] Y. Dong, X. Zhao, and X. Huang, "Dependability analysis of deep reinforcement learning based robotics and autonomous systems through probabilistic model checking," in *2022 IEEE/RSJ international conference on intelligent robots and systems*, 2022.

[118] X. Sun, H. Khedr, and Y. Shoukry, "Formal verification of neural network controlled autonomous systems," in *Proceedings of the 22nd ACM international conference on hybrid systems: Computation and control*, 2019.

[119] R. Hawkins, C. Paterson, C. Picardi, Y. Jia, *et al.*, “Guidance on the assurance of machine learning in autonomous systems (AMLAS),” arXiv preprint arXiv:2102.01564, 2021.

[120] C. Picardi, R. Hawkins, C. Paterson, and I. Habli, “Transfer assurance for machine learning in autonomous systems,” *SafeAI*, 2023.

[121] L. Brunke, M. Greeff, A. Hall, Z. Yuan, *et al.*, “Safe learning in robotics: From learning-based control to safe reinforcement learning,” *Annual Review of Control, Robotics, and Autonomous Systems*, 2022.

[122] B. Lütjens, M. Everett, and J. How, “Safe reinforcement learning with model uncertainty estimates,” in *2019 international conference on robotics and automation*, 2019.

[123] H. Taheri, S. Hosseini, and M. Nekoui, “Deep reinforcement learning with enhanced ppo for safe mobile robot navigation,” arXiv preprint arXiv:2405.16266, 2024.

[124] H. Cao, Y. Mao, L. Sha, and M. Caccamo, “Physics-regulated deep reinforcement learning: Invariant embeddings,” in *International conference on learning representations*, 2024.

[125] M. Tayal, A. Singh, S. Kolathaya, and S. Bansal, “A physics-informed machine learning framework for safe and optimal control of autonomous systems,” arXiv preprint arXiv:2502.11057, 2025.

[126] H. Ahmad, E. Sabouni, A. Wasilkoff, P. Budhraja, *et al.*, “Hierarchical multi-agent reinforcement learning with control barrier functions for safety-critical autonomous systems,” arXiv preprint arXiv:2507.14850, 2025.

[127] C. Huang, S. Xu, Z. Wang, S. Lan, W. Li, *et al.*, “Opportunistic intermittent control with safety guarantees for autonomous systems,” in *ACM/IEEE international conference on cyber-physical systems*, 2020.

[128] X. Pan *et al.*, “Characterizing attacks on deep reinforcement learning,” arXiv preprint arXiv:1907.09470, 2019.

[129] C. Yang, J. Qi, P. Chen, Y. Ouyang, *et al.*, “Enhanced adversarial strategically-timed attacks against deep reinforcement learning,” *ICASSP*, 2020.

[130] S. Feng *et al.*, “Dense reinforcement learning for safety validation of autonomous vehicles,” *Nature*, 2023.

[131] H. Niu, Z. Ji, F. Arvin, B. Lennox, H. Yin, *et al.*, “Accelerated sim-to-real deep reinforcement learning: Learning collision avoidance from human player,” in *IEEE/SICE international symposium on system integration*, 2021.

[132] V. Goecks, G. Gremillion, V. Lawhern, *et al.*, “Efficiently combining human demonstrations and interventions for safe training of autonomous systems in real-time,” in *Proceedings of the AAAI conference on artificial intelligence*, 2019.

[133] L. Xie, S. Wang, S. Rosa, A. Markham, *et al.*, “Learning with training wheels: Speeding up training with a simple controller for deep reinforcement learning,” in *2018 IEEE international conference on robotics and automation*, 2018.

[134] H. Janardhanan, “Hierarchical reinforcement learning for scalable autonomous systems: A multi-level control approach,” in *2025 8th international conference on circuit, system and signal processing*, 2025.

[135] K. Shaheen, M. Hanif, O. Hasan, and M. Shafique, “Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks,” *Journal of Intelligent & Robotic Systems*, 2022.

[136] J. Rosero, N. Cardozo, *et al.*, “Multi-objective deep reinforcement learning optimisation in autonomous systems,” in *IEEE international conference on self-organizing systems*, 2024.

[137] T. Le, B. Hung, and P. V. Huy, “Search-based planning and reinforcement learning for autonomous systems and robotics,” *Deep Learning for Unmanned Systems*, 2021.

[138] C. Hoel, K. Driggs-Campbell, K. Wolff, *et al.*, “Combining planning and deep reinforcement learning in tactical decision making for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, 2019.

[139] T. Huynh and K. Nguyen, “Integrative AI framework for robotics: LLM-enabled reinforcement learning in object manipulation and task planning,” *Robotics and Autonomous Systems*, 2025.

[140] A. Shiri, M. Navardi, T. Manjunath, N. Waytowich, *et al.*, “Efficient language-guided reinforcement learning for resource-constrained autonomous systems,” *IEEE Micro*, 2022.

[141] S. Jain, “Beyond traditional algorithms: Harnessing reinforcement learning and generative AI for next-generation autonomous systems,” *ICONIC Research and Engineering Journals*, 2019.

[142] A. Sehgal, H. La, S. Louis, *et al.*, “Deep reinforcement learning using genetic algorithm for parameter optimization,” in *2019 third IEEE international conference on robotic computing*, 2019.

[143] T. Kobayashi, “Adaptive and multiple time-scale eligibility traces for online deep reinforcement learning,” *Robotics and Autonomous Systems*, 2022.

[144] K. Rana, V. Dasagi, J. Haviland, *et al.*, “Bayesian controller fusion: Leveraging control priors in deep reinforcement learning for robotics,” *The International Journal of Robotics Research*, 2023.

[145] A. Morgan, D. Nandha, G. Chalvatzaki, *et al.*, “Model predictive actor-critic: Accelerating robot skill acquisition with deep reinforcement learning,” in *IEEE international conference on robotics and automation*, 2021.

[146] B. Lian, Y. Kartal, F. Lewis, D. Mikulski, *et al.*, "Anomaly detection and correction of optimizing autonomous systems with inverse reinforcement learning," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.

[147] V. Talpaert, I. Sobh, B. Kiran, P. Mannion, *et al.*, "Exploring applications of deep reinforcement learning for real-world autonomous driving systems," arXiv preprint arXiv:1901.01536, 2019.

[148] A. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," arXiv preprint arXiv:1704.02532, 2017.

[149] S. Nageshrao, H. Tseng, and D. Filev, "Autonomous highway driving using deep reinforcement learning," in *IEEE international conference on systems, man and cybernetics*, 2019.

[150] J. Chen, S. Li, and M. Tomizuka, "Interpretable end-to-end urban autonomous driving with latent deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[151] J. Lu, G. Alcan, and V. Kyrki, "Integrating expert guidance for efficient learning of safe overtaking in autonomous driving using deep reinforcement learning," arXiv preprint arXiv:2308.09456, 2023.

[152] D. Isele, R. Rahimi, A. Cosgun, *et al.*, "Navigating occluded intersections with autonomous vehicles using deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2018.

[153] B. Elallid, H. Alaoui, and N. Benamar, "Deep reinforcement learning for autonomous vehicle intersection navigation," arXiv preprint arXiv:2310.08595, 2023.

[154] J. Hossain, "Autonomous driving with deep reinforcement learning in carla simulation," arXiv preprint arXiv:2306.11217, 2023.

[155] M. O'Kelly, H. Zheng, D. Karthik, *et al.*, "F1tenth: An open-source evaluation environment for continuous control and reinforcement learning," in *Proceedings of machine learning research*, 2020.

[156] M. Spryn, A. Sharma, D. Parkar, and M. Shrimal, "Distributed deep reinforcement learning on the cloud for autonomous driving," in *Proceedings of the 1st ACM SIGSOFT international workshop on machine learning for software engineering*, 2018.

[157] S. Lu, L. He, S. Li, Y. Luo, J. Wang, *et al.*, "Hierarchical end-to-end autonomous driving: Integrating BEV perception with deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2025.

[158] M. Huegle, G. Kalweit, B. Mirchevska, *et al.*, "Dynamic input for deep reinforcement learning in autonomous driving," in *IEEE/RSJ international conference on intelligent robots and systems*, 2019.

[159] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Robotics and Autonomous Systems*, 2019.

[160] X. Ma, J. Li, M. Kochenderfer, D. Isele, *et al.*, "Reinforcement learning for autonomous driving with latent state inference and spatial-temporal relationships," in *IEEE/RSJ international conference on intelligent robots and systems*, 2021.

[161] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, *et al.*, "Champion-level drone racing using deep reinforcement learning," *Nature*, 2023.

[162] Y. Song, M. Steinweg, E. Kaufmann, *et al.*, "Autonomous drone racing with deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2021.

[163] B. Evans, H. Engelbrecht, *et al.*, "High-speed autonomous racing using trajectory-aided deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2023.

[164] N. Kunda, P. Kc, M. Pandey, and A. Kumaar, "Reward design and hyperparameter tuning for generalizable deep reinforcement learning agents in autonomous racing," *Scientific Reports*, 2025.

[165] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, *et al.*, "How to train your robot with deep reinforcement learning: Lessons we have learned," *The International Journal of Robotics Research*, 2021.

[166] O. Arriaga, M. Valdenegro-Toro, M. Muthuraja, *et al.*, "Perception for autonomous systems (paz)," arXiv preprint arXiv:2010.14541, 2020.

[167] Z. Li, A. Samanta, Y. Li, A. Soltoggio, *et al.*, ": On-device real-time deep reinforcement learning for autonomous robotics," in *IEEE real-time systems symposium*, 2023.

[168] M. Kim, J. Kim, M. Jung, and H. Oh, "Towards monocular vision-based autonomous flight through deep reinforcement learning," *Expert Systems with Applications*, 2022.

[169] G. Ciabatti, S. Daftry, *et al.*, "Autonomous planetary landing via deep reinforcement learning and transfer learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2021.

[170] A. Rodriguez-Ramos, C. Sampedro, H. Bavle, *et al.*, "A deep reinforcement learning strategy for UAV autonomous landing on a moving platform," *Journal of Intelligent & Robotic Systems*, 2019.

[171] Z. Jiang and G. Song, "A deep reinforcement learning strategy for UAV autonomous landing on a platform," in *2022 international conference on unmanned aircraft systems*, 2022.

[172] K. Kang, S. Belkhale, G. Kahn, P. Abbeel, *et al.*, "Generalization through simulation: Integrating simulated and real data into deep reinforcement learning for vision-based autonomous flight," in *International conference on robotics and automation*, 2019.

[173] J. Westheider, J. Rückin, *et al.*, "Multi-UAV adaptive path planning using deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2023.

[174] A. Kolling, V. Kich, J. de Jesus, *et al.*, "Parallel distributional prioritized deep reinforcement learning for unmanned aerial vehicles," in *Latin american robotics symposium and brazilian symposium on robotics*, 2023.

[175] T. Hickling, M. Hogan, A. Tammam, and N. Aouf, "Deep reinforcement learning based autonomous decision-making for cooperative uavs: A search and rescue real world application," arXiv preprint arXiv:2502.20326, 2025.

[176] Y. Flato, R. Harel, A. Tamar, R. Nathan, *et al.*, "Revealing principles of autonomous thermal soaring in windy conditions using vulture-inspired deep reinforcement-learning," *Nature Communications*, 2024.

[177] G. Shetty, M. Ramezani, H. Habibi, H. Voos, *et al.*, "Motion control in multi-rotor aerial robots using deep reinforcement learning," in *International conference on robotics and automation*, 2025.

[178] T. Hosmer and P. Tigas, "Deep reinforcement learning for autonomous robotic tensegrity (ART)," in *Proceedings of the 39th annual conference of the association for computer aided design in architecture*, 2019.

[179] K. Wu, H. Wang, M. Esfahani, *et al.*, "BND*-DDQN: Learn to steer autonomously through deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[180] I. Carlucho, M. D. Paula, S. Wang, *et al.*, "AUV position tracking control using end-to-end deep reinforcement learning," *OCEANS*, 2018.

[181] Q. Zhang, J. Lin, Q. Sha, B. He, and G. Li, "Deep interactive reinforcement learning for path following of autonomous underwater vehicle," *Ieee Access*, 2020.

[182] W. Wang, X. Cao, A. Gonzalez-Garcia, L. Yin, *et al.*, "Deep reinforcement learning based tracking control of an autonomous surface vessel in natural waters," arXiv preprint arXiv:2302.08100, 2023.

[183] E. Meyer, A. Heiberg, A. Rasheed, and O. San, "COLREG-compliant collision avoidance for unmanned surface vehicle using deep reinforcement learning," *Ieee Access*, 2020.

[184] S. Reddy, A. Dragan, and S. Levine, "Shared autonomy via deep reinforcement learning," arXiv preprint arXiv:1802.01744, 2018.

[185] Y. Chen, M. Everett, M. Liu, *et al.*, "Socially aware motion planning with deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2017.

[186] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2019.

[187] Y. Gao, F. Yang, M. Frisk, D. Hemandez, *et al.*, "Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning," in *2019 28th IEEE international conference on robot and human interactive communication*, 2019.

[188] M. Everett, Y. Chen, and J. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *Ieee Access*, 2021.

[189] A. Ghadirzadeh, X. Chen, W. Yin, Z. Yi, *et al.*, "Human-centered collaborative robots with deep reinforcement learning," *IEEE Robotics and Automation Letters*, 2020.

[190] N. Waytowich, V. Goecks, and V. Lawhern, "Cycle-of-learning for autonomous systems from human interaction," arXiv preprint arXiv:1808.09572, 2018.

[191] H. Mahmoud and I. Ibrahim, "Adaptive hybrid algorithms for real-time decision-making in autonomous systems," *Asian Journal of Research in Computer Science*, 2025.

[192] J. Cai, A. Du, X. Liang, and S. Li, "Prediction-based path planning for safe and efficient human–robot collaboration in construction via deep reinforcement learning," *Journal of computing in civil engineering*, 2023.

[193] M. Kasaei, M. Abreu, N. Lau, A. Pereira, *et al.*, "Robust biped locomotion using deep reinforcement learning on top of an analytical control approach," *Robotics and Autonomous Systems*, 2021.

[194] Z. Xie, G. Berseth, P. Clary, J. Hurst, *et al.*, "Feedback control for cassie with deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2018.

[195] S. Kim, M. Sorokin, J. Lee, and S. Ha, "Humanconquad: Human motion control of quadrupedal robots using deep reinforcement learning," *SIGGRAPH Asia*, 2022.

[196] S. Ha, J. Kim, and K. Yamane, "Automated deep reinforcement learning environment for hardware of a modular legged robot," in *2018 15th international conference on ubiquitous robots*, 2018.

[197] A. Lele, Y. Fang, J. Ting, *et al.*, "Learning to walk: Spike based reinforcement learning for hexapod robot central pattern generation," in *2020 2nd IEEE international conference on artificial intelligence circuits and systems*, 2020.

[198] R. Zhu, D. Zhang, and B. Lo, "Deep reinforcement learning based semi-autonomous control for robotic surgery," arXiv preprint arXiv:2204.05433, 2022.

[199] K. Li *et al.*, "Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning," in *IEEE international conference on robotics and automation*, 2021.

[200] H. Hase, M. Azampour, M. Tirindelli, *et al.*, "Ultrasound-guided robotic navigation with deep reinforcement learning," in *IEEE/RSJ international conference on intelligent robots and systems*, 2020.

[201] M. Macaulay and M. Shafiee, "Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure," *Autonomous Intelligent Systems*, 2022.

[202] B. Felbrich, T. Schork, and A. Menges, "Autonomous robotic additive manufacturing through distributed model-free deep reinforcement learning in computational design environments," *Construction Robotics*, 2022.

[203] A. Zhu, T. Dai, G. Xu, P. Pauwels, *et al.*, "Deep reinforcement learning for real-time assembly planning in robot-based prefabricated construction," *IEEE Transactions on Automation Science and Engineering*, 2023.

[204] O. Azulay and A. Shapiro, "Wheel loader scooping controller using deep reinforcement learning," *IEEE access*, 2021.

[205] O. Yapar, "Reinforcement learning in autonomous defense systems: Strategic applications and challenges," *World Journal of Advanced Engineering Technology and Sciences*, 2024.

[206] H. Hohenfeld, D. Heimann, F. Wiebe, and F. Kirchner, "Quantum deep reinforcement learning for robot navigation tasks," *IEEE Access*, 2024.

[207] S. Tomar, S. Alam, S. Kumar, and A. Mathur, "Quantum-enhanced hybrid reinforcement learning framework for dynamic path planning in autonomous systems," arXiv preprint arXiv:2504.20660, 2025.

[208] L. Zanatta, F. Barchi, S. Manoni, S. Tolu, A. Bartolini, *et al.*, "Exploring spiking neural networks for deep reinforcement learning in robotic tasks," *Scientific Reports*, 2024.

[209] Y. Kadokawa, T. Kodera, Y. Tsurumine, *et al.*, "Robust iterative value conversion: Deep reinforcement learning for neurochip-driven edge robots," *Robotics and Autonomous Systems*, 2024.

[210] L. Lei, Y. Tan, K. Zheng, S. Liu, K. Zhang, *et al.*, "Deep reinforcement learning for autonomous internet of things: Model, applications and challenges," *IEEE Communications Surveys & Tutorials*, 2020.

[211] S. Ma, J. Ruan, Y. Du, R. Bucknall, *et al.*, "An end-to-end deep reinforcement learning based modular task allocation framework for autonomous mobile systems," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.

[212] S. Vinay, "Applications of neurocomputing in autonomous systems and robotics," *International Journal of Neurocomputing (IJN)*, 2024.