ISSN: 3008-1394

SOCIALLY-AWARE ROBOT NAVIGATION VIA DEEP REINFORCEMENT LEARNING: A CRITICAL REVIEW AND FUTURE DIRECTIONS

Qichao Jiang[1*], Hazirah Bee Yusof Ali and [2]Li Ting[3]

[1, 2, 3*]City University Malaysia

[1*]316406024@qq.com, [2] hazirah.bee@city.edu.my, [3] 284170024@qq.com

Abstract

The integration of Autonomous Mobile Robots (AMRs) into human-centric environments presents a formidable scientific challenge: achieving navigation that is not only safe and efficient but also socially compliant. Traditional path planning algorithms, designed for structured and static worlds, fundamentally fail to address the complex, interactive, and socially governed dynamics of human crowds. Deep Reinforcement Learning (DRL) has emerged as a powerful paradigm for learning adaptive navigation policies through continuous interaction. This paper provides a comprehensive and critical review of the state-of-the-art in DRL for socially-aware robot navigation. We begin by framing the core problem as a governing trilemma among safety, efficiency, and social compliance. We then critically analyze the inherent limitations of classical navigation paradigms in social contexts, specifically focusing on the "Freezing Robot Problem." The core of this review is a deep dive into the thematic debates shaping modern DRL-based approaches, including architectural philosophies, reward engineering across physics-based and psychology-based domains, and sim-to-real transfer strategies. Finally, we outline a roadmap for next-generation social navigation, emphasizing Large Language Model (LLM) integration and explainable AI.

Keywords

Autonomous Mobile Robots, Deep Reinforcement Learning, Human-Robot Interaction, Social Compliance, Socially-Aware Navigation.

1. Introduction

The integration of intelligent systems into the fabric of human society is a key manifestation of the Fourth Industrial Revolution. Autonomous Mobile Robots (AMRs) have evolved from laboratory curiosities into indispensable cornerstones in logistics, healthcare, and construction(Attalla et al., 2023; Fontani et al., 2025; Keith & La, 2024). The global AMR market is projected to exceed $50 billion by 2030 (Lässig et al., 2021).yet their transition to unstructured human-centric spaces—such as bustling airport terminals or clinical corridors—is predicated on solving the intricate problem of socially-aware navigation.

ISSN: 3008-1394

Unlike static obstacles, humans are rational agents whose behavior is driven by opaque internal states such as intentions and emotions (Selvaggio et al., 2021). A robot navigating this landscape must act in a manner that is not just collision-free, but also socially acceptable and predictable (Faria et al., 2021; Möller et al., 2021).As research consistently indicates, this challenge is best framed as a governing trilemma (see Fig. 1) :

- Safety: Preventing physical harm and respecting subjective human comfort (Hetherington et al., 2021; Standardization, 2014).
- Efficiency: Optimizing task accomplishment through metrics like time-to-goal and energy consumption (Verstraete & Muhammad, 2024).
- Social Compliance: Adhering to implicit conventions, such as respecting personal space (proxemics) and moving legibly (Guillén-Ruiz et al., 2023; Tao et al., 2025, Early Access).

Figure 1 The governing trilemma of socially-aware navigation.

The tension between these objectives raises a fundamental question: How can a robot learn to navigate these trade-offs as intuitively as a human pedestrian?(Schulz, 2021). This paper provides a critical review of DRL's application to this trilemma, synthesizing the state-of-the-art, identifying key research gaps, and charting a course for future inquiry.

2. Critical Evaluation Of Classical Navigation Paradigms

The shift toward Deep Reinforcement Learning is largely a response to the inherent brittleness of traditional navigation logic when faced with social complexity.

2.1 Traditional Path Planning (A* and RRT*)

Historically, algorithms like A* (Li et al., 2023) and RRT* (Abdel-Jaber et al., 2022) have been the gold standard for global path planning (Tao et al., 2025). However, these methods typically rely on static occupancy grids. In dynamic

ISSN: 3008-1394

social settings—where human movement is fluid and unpredictable—the latency inherent in re-planning often renders these global paths obsolete the moment they are generated.

2.2 Reactive and Reciprocal Methods (APF and ORCA)

Local planners (or reactive layers) use real-time sensor data to avoid immediate obstacles. However, their core assumptions break down in dense, interactive crowds.

Table 1 Comparison Of Traditional Path Planning Algorithms.

Algorithm Class	Representative	Core Limitation in Social Context
Graph-Search	A*, RRT*	Static snapshots; fails to model human-robot interaction dynamics.
Reactive	APF (Potential Fields)	Prone to local minima; "oscillatory" behavior in dense crowds.
Reciprocal	ORCA	Prone to local minima; "oscillatory" behavior in dense crowds.
Predictive	DWA	Myopic; leads to the "Freezing Robot Problem" when all paths seem blocked.

- Artificial Potential Field (APF): This method treats the robot as a particle in a force field. While computationally efficient, it is notorious for falling into local minima and exhibiting oscillatory behavior in dense crowds (Khatib, 1986).

- Optimal Reciprocal Collision Avoidance (ORCA): This method improved this by assuming "reciprocity"—the idea that both the robot and the human will take equal responsibility for avoiding a collision. However, in the real world, humans often exhibit non-cooperative behaviors, expecting the robot to yield exclusively. This mismatch leads to stalled motion or unsafe proximity(Van Den Berg et al., 2011).

2.3 The "Freezing Robot Problem" (FRP)

A critical failure occurs when a robot`s local planner (like DWA) determines that all possible trajectories will eventually lead to a collision due to the high density of moving humans. The robot then decides the safest action is to stop completely, effectively "freezing" and becoming an obstacle itself.

In essence, classical methods fail because they are myopic in decision-making, naive in modeling humans, and consequently unacceptable in their emergent

ISSN: 3008-1394

social behavior. They lack the ability to learn from experience, understand social context, or anticipate the long-term evolution of a scene.

3.Deep Reinforcement Learning: The New Frontier

Deep Reinforcement Learning (DRL) reframes navigation as a Markov Decision Process (MDP), enabling agents to internalize complex social trade-offs through cumulative experience.

3.1 Architectural Philosophies

The choice of DRL algorithm dictates how the agent learns. The main families are summarized in Table .

- State Space (S): A high-dimensional vector combining exteroceptive sensor data (e.g., 360° LiDAR scans), proprioceptive robot state (e.g., velocity), and task information (e.g., distance and angle to the goal).
- Action Space (A): Typically a continuous 2D vector representing linear and angular velocity commands, enabling smooth control.
- Reward Function (R): The most critical component for shaping behavior. It is engineered as a multi-objective function that encodes the safety-efficiency-social compliance trilemma.

Transition Probability (P): The environment's dynamics, which are unknown and include the complex, unwritten rules of human behavior. DRL methods are typically model-free, learning a policy without explicitly modeling P.

Table 2 Comparative Anylysis Of Mainstream DRL Algorithms

Algorithm	**Category**	**Action Space**	**Core Advantages**	**Key Challenges**
DQN	Value-Based	Discrete	Pioneering work for high-dim state inputs; stable and effective.	Cannot handle continuous actions directly; Q-value overestimation.
PPO	Policy-Based	Discrete /Continuous	Simple to implement; enhances training stability by clipping policy updates.	Relatively lower sample efficiency.

ISSN: 3008-1394

DDPG	Actor-Critic	Continuous	Successfully extended DRL to continuous control domains.	Sensitive to hyperparameters; unstable training; Q-value overestimation.
TD3	Actor-Critic	Continuous	Effectively mitigates DDPG's Q-value overestimation; more stable.	Algorithm is relatively more complex.
SAC	Actor-Critic	Continuous	Encourages exploration via max entropy; high sample efficiency and stability.	Theory is relatively complex.

3.2 Leading DRL Algorithms

The DRL landscape is dominated by three families of algorithms, summarized in Table II.

- Value-Based Methods (e.g., DQN): Learn an action-value function Q(s,a) and derive the policy by choosing the action with the highest value. They excel in discrete action spaces but struggle with continuous control (Li et al., 2023).

- Policy-Based Methods (e.g., PPO): Directly learn a stochastic policy π(a|s). They are well-suited for continuous action spaces and often exhibit more stable convergence properties (Attalla et al., 2023).

- Actor-Critic Methods (e.g., SAC, TD3): A hybrid approach that learns both a policy (the Actor) and a value function (the Critic). The Critic evaluates the Actor's actions, providing low-variance learning signals. This paradigm, especially with algorithms like Soft Actor-Critic (SAC), represents the state-of-the-art for continuous control due to its high sample efficiency and robustness.

3.3 Reward Engineering: Encoding Social Norms

The behavior of a DRL agent is fundamentally defined by its reward function, typically structured as:

$$R_{total} = w_g R_{goal} + w_s R_{safety} + w_p R_{proxemics} + w_c R_{smoothness}$$

- Proxemics Rewards: Penalizing the robot for entering the "intimate zone" (0.45m) or "personal zone" (1.2m) of a human.

ISSN: 3008-1394

- Interaction Rewards: Rewarding the robot for maintaining a consistent speed when passing others, preventing abrupt, frightening movements.

- Goal Progress: Ensuring efficiency is not sacrificed for excessive caution.

Critically, the weights w in current research are primarily assigned through heuristic manual tuning. This trial-and-error process highlights a significant bottleneck, suggesting that future systems must transition to Inverse Reinforcement Learning (IRL) to automatically infer human-like weight distributions from demonstration data.

4.Bridging The Sim-to-real Gap

The application of DRL to social navigation is a vibrant research area defined by several key thematic debates on the best architectural and methodological approaches.

4.1 Architectural Philosophies: End-to-End vs. Modular Design

While End-to-End learning offers seamless mapping from sensors to actions, it suffers from a lack of interpretability. More importantly, these monolithic networks are prone to Catastrophic Forgetting—where a model trained in low-density simulations loses its foundational navigation capabilities when exposed to high-density social noise or novel edge cases. Modular designs mitigate this by isolating perception from policy, allowing for safer verification.

- End-to-End Learning: This philosophy advocates for a single, monolithic neural network that maps raw sensor data directly to motor commands. The allure is that the network can discover its own optimal intermediate representations, bypassing potentially suboptimal human-engineered pipelines. However, this "black box" approach suffers from poor interpretability, data inefficiency, and challenges in safety verification.

- Modular Design: This pragmatic engineering approach decomposes the problem into a pipeline of specialized modules (e.g., Perception → Prediction → Planning → Control). This enhances interpretability, allows for the integration of pre-existing knowledge (e.g., using a pre-trained object detector), and facilitates safety verification through dedicated safety layers. The state-of-the-art is converging towards hybrid systems that leverage deep learning for complex modules like perception and prediction, while retaining modular separation for planning and safety

4.2 Reward Function Engineering: A Spectrum of Approaches

Designing a reward function that captures social norms is a central challenge. As summarized in Table 3, approaches can be categorized by their source of knowledge.

ISSN: 3008-1394

- Physics-Based (SFM): Draws inspiration from the Social Force Model (Fontani et al., 2025), treating social interaction as a set of attractive and repulsive forces. This is intuitive but can oversimplify complex social rules.

- Psychology-Based (Proxemics): Grounds rewards in the theory of proxemics (personal space) (Campagna & Rehm, 2025), penalizing intrusions into culturally-defined social zones. This is more human-centric but can be rigid and context-dependent.

- Data-Driven (Learning from Demonstration): Uses expert human demonstrations to either infer the underlying reward function (Inverse Reinforcement Learning - IRL) or learn the policy directly (Imitation Learning - IL). This can capture nuanced behaviors but is limited by the quality and coverage of the expert data.

Table 2 Comparision Of Reward Design Philosophes

Approach	Source of Knowledge	Primary Engineering Effort	Generalization Power	Interpretability	Key Limitation
SFM-Based	Physics Analogy	Manual tuning of force parameters and weights.	Limited by the model's physical assumptions.	High (Forces are intuitive).	Oversimplification of social rules.
Proxemics-Based	Social Psychology	Manual design of zone boundaries and penalties.	Culturally-bound and context-dependent.	Medium (Zones are clear, but effects are non-linear).	Rigidity; ignores orientation and context.
LfD-Based (IRL/IL)	Expert Data	Collection and curation of a large, diverse dataset.	Limited by the coverage and quality of the demonstration data.	Low (Learned reward/policy is a black box).	Suboptimality of expert; data cost.

4.3 The Simulation-to-Reality (Sim-to-Real) Challenge

The immense sample complexity of DRL necessitates training in simulation. However, the "reality gap" between the simulator and the physical world is a major obstacle. Key strategies to bridge this gap include:

- Domain Randomization (DR):Randomizing physical parameters (friction, sensor noise) and social behaviors (pedestrian speeds and aggressiveness)

ISSN: 3008-1394

during training to ensure the policy generalizes to real-world hallways. (Guillén-Ruiz et al., 2023).

- Domain Adaptation & System Identification: These techniques use a small amount of real-world data to either learn a mapping between the simulated and real domains or to identify the true physical parameters of the robot to create a more accurate "digital twin" simulation.

Figure 2. The Sim2Real Gap Challenge.

5.Future Research Directions

Synthesizing the limitations of current work reveals several promising avenues for future research.

- Human Intent Prediction as a First-Class Citizen: The most critical gap is the reactive nature of current planners. Future architectures must move towards a proactive, predictive paradigm by integrating dedicated, explicit human intent prediction modules (e.g., using Transformer networks) into the DRL loop. The planner's input should not be where people are, but where they are predicted to be.

- Context-Aware Reward Modulation: To overcome the rigidity of monolithic reward functions, future work should explore mechanisms for context-aware reward modulation. This would allow a robot to dynamically adjust its priorities (e.g., weighting efficiency higher in an empty hall and social compliance higher in a dense crowd) based on a high-level perception of the social context.

- Hybrid Architectures for Proactive and Collaborative Planning: The future lies in hybrid systems that combine global planners with local DRL agents that are both predictive and collaborative, as shown in Fig. 3.

- LLM-Augmented Social Reasoning: Using Large Language Models to provide "high-level social common sense" (e.g., distinguishing between a formal meeting and a casual lobby) to modulate DRL reward weights dynamically.

ISSN: 3008-1394

- Interpretable and Verifiable Social DRL: Integrating principles from Explainable AI (XAI) and formal methods is crucial for building trustworthy systems. This involves designing architectures that are more transparent and incorporating safety kernels that can provide hard guarantees on collision avoidance.

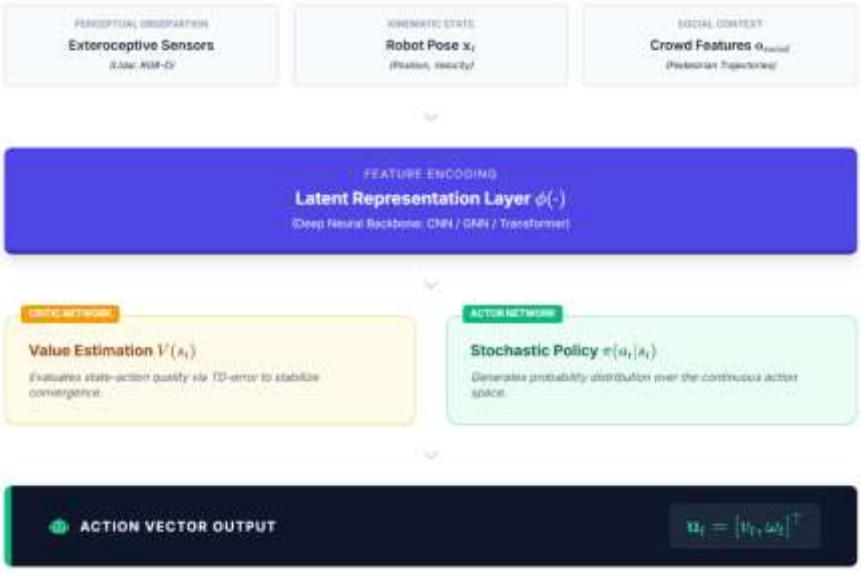

Figure 3. A Typical DRL Agent Architecture for Navigation.

6. Conclusion

This review has synthesized the current state of DRL-based social navigation, highlighting the shift from collision avoidance to social compliance. While DRL offers unparalleled adaptability, the challenge remains in balancing the navigation trilemma in high-density, dynamic environments.

Acknowledgement

The authors would like to thank City University Malaysia for the resources provided during this research.

Funding

The author(s) received no specific funding for this work.

Author Contribution

Author1 prepared the literature review . Author2 oversaw the article writing.

Conflict of Interest

The authors have no conflicts of interest to declare.

ISSN: 3008-1394

References

[1] Abdel-Jaber, H., Devassy, D., Al Salam, A., Hidaytallah, L., & El-Amir, M. (2022). A review of deep learning algorithms and their applications in healthcare. Algorithms, 15(2), 71.

[2] Attalla, A., Attalla, O., Moussa, A., Shafique, D., Raean, S. B., & Hegazy, T. (2023). Construction robotics: Review of intelligent features. International Journal of Intelligent Robotics and Applications, 7(3), 535–555. https://doi.org/10.1007/s41315-023-00275-1

[3] Campagna, G., & Rehm, M. (2025). A Systematic Review of Trust Assessments in Human–Robot Interaction. ACM Transactions on Human-Robot Interaction, 14(2), 1–35. https://doi.org/10.1145/3706123

[4] Faria, M., Melo, F. S., & Paiva, A. (2021). Understanding robots: Making robots more legible in multi-party interactions. 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), 1031–1036. https://ieeexplore.ieee.org/abstract/document/9515485/

[5] Fontani, M., Luglio, S. M., Gagliardi, L., Peruzzi, A., Frasconi, C., Raffaelli, M., & Fontanelli, M. (2025). A systematic review of 59 field robots for agricultural tasks: Applications, trends, and future directions. Agronomy, 15(9), 2185.

[6] Guillén-Ruiz, S., Bandera, J. P., Hidalgo-Paniagua, A., & Bandera, A. (2023). Evolution of socially-aware robot navigation. Electronics, 12(7), 1570.

[7] Hetherington, N. J., Croft, E. A., & Van der Loos, H. M. (2021). Hey robot, which way are you going? Nonverbal motion legibility cues for human-robot spatial interaction. IEEE Robotics and Automation Letters, 6(3), 5010–5015.

[8] Keith, R., & La, H. M. (2024). Review of autonomous mobile robots for the warehouse environment (arXiv:2406.08333). arXiv. https://doi.org/10.48550/arXiv.2406.08333

[9] Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. The International Journal of Robotics Research, 5(1), 90–98. https://doi.org/10.1177/027836498600500106

[10] Lässig, R., Lorenz, M., Sissimatos, E., Wicker, I., & Buchner, T. (2021). Robotics outlook 2030: How intelligence and mobility will shape the future. Boston Consulting Group, 28. https://arabesque.ch/wp-content/uploads/2022/09/BCG-Study-shape-of-the-robotics-industry.pdf

ISSN: 3008-1394

[11] Li, L., Chen, Y., Sun, F., & Guo, R. (2023). Robot navigation using reinforcement learning with multi attention fusion in crowd. In F. Sun, J. Li, H. Liu, & Z. Chu (Eds.), Cognitive Computation and Systems (Vol. 1732, pp. 247–258). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2789-0_21

[12] Möller, R., Furnari, A., Battiato, S., Härmä, A., & Farinella, G. M. (2021). A survey on human-aware robot navigation. Robotics and Autonomous Systems, 145, 103837.

[13] Schulz, V. H. (2021). Book reviews. SIAM Review, 63(2), 419–431. https://doi.org/10.1137/21N975254

[14] Selvaggio, M., Cognetti, M., Nikolaidis, S., Ivaldi, S., & Siciliano, B. (2021). Autonomy in physical human-robot interaction: A brief survey. IEEE Robotics and Automation Letters, 6(4), 7989–7996.

[15] Standardization, I. O. for. (2014). Robots and robotic devices: Safety requirements for personal care robots.

[16] Tao, X., Li, H., Chen, Z., & Xu, D. (2025). Robot navigation in dynamic and crowded environments. IEEE Transactions on Control Systems Technology. https://ieeexplore.ieee.org/abstract/document/11030716/.(Early Access)

[17] Van Den Berg, J., Guy, S. J., Lin, M., & Manocha, D. (2011). Reciprocal n-body collision avoidance Robot. Research, 3–19.

[18] Verstraete, T., & Muhammad, N. (2024). Pedestrian collision avoidance in autonomous vehicles: A review. Computers, 13(3), 78.