ISSN: 3008-1394

ADVANCES IN DEEP REINFORCEMENT LEARNING FOR AUTONOMOUS MOBILE ROBOT NAVIGATION: A SYSTEMATIC REVIEW OF ALGORITHMS, SIM2REAL, AND SAFETY

Li Ting*

*City University Malaysia

*284170024@qq.com,

Hazirah Bee Yusof Ali[2]

City University Malaysia

hazirah.bee@city.edu.my

Jiang Qichao

City University Malaysia

316406024@qq.com

***Corresponding author:** *284170024@qq.com

Abstract

Despite decades of progress, navigating autonomous mobile robots (AMRs) through stochastic and non-deterministic environments remains a formidable challenge. While classical geometric-based planners provide rigorous theoretical guarantees, their performance often degrades under high-dimensional uncertainty. This review dissects the paradigm shift toward Deep Reinforcement Learning (DRL), emphasizing its capacity for end-to-end perception-to-action mapping. We categorize contemporary DRL architectures into value-based, policy-gradient, and actor-critic lineages, evaluating their efficiency in complex workspaces (Haarnoja et al., 2018; Hasselt et al., 2015; Schulman et al., 2017). Crucially, we scrutinize the field's persistent "pain points": training sample inefficiency, the notorious Sim2Real gap (Da et al., 2025; Loquercio, 2023), and the exigency for verifiable safety constraints (Gu et al., 2024). By synthesizing landmark studies from 2015 to 2025, this paper contributes a novel taxonomy and explores the emerging trajectory of Foundation Model-driven navigation, providing a roadmap for the next generation of socially-aware and resilient AMR systems.

Keywords

Autonomous Mobile Robots, Deep Reinforcement Learning, Sim2Real, Safety-Critical Navigation, Foundation Models, Industry 4.0.

1.Introduction

Industry 4.0 has thrust AMRs into the spotlight, demanding they operate not just in static factories, but in the messy, unstructured reality of public spaces (Anwer et al., 2024). Historically, path planning relied on neat geometric abstractions (Lin et al., 2022). However,

ISSN: 3008-1394

the modern consensus is shifting: DRL offers a more "visceral" approach to navigation by merging the raw perceptual power of Deep Neural Networks (DNNs) with the goal-oriented logic of reinforcement learning. As depicted in Fig. 1, the DRL agent learns through trial-and-error—a process akin to biological learning. This avoids the brittle nature of predefined maps, which is vital for swerving through unpredictable pedestrian traffic (Singh et al., 2023). Nevertheless, DRL is no silver bullet. Researchers still grapple with agonizingly slow convergence times and the "black box" nature of reward design. For instance, a basic DQN might spin in circles for weeks before identifying an optimal path (K. Zhu & Zhang, 2021), and even then, a model that performs flawlessly in a simulator may fail instantly when faced with real-world sensor noise (Huber et al., 2024; Pitkevich & Makarov, 2024). This review aims to cut through the hype and analyze the concrete breakthroughs that will define the future of robot intelligence.

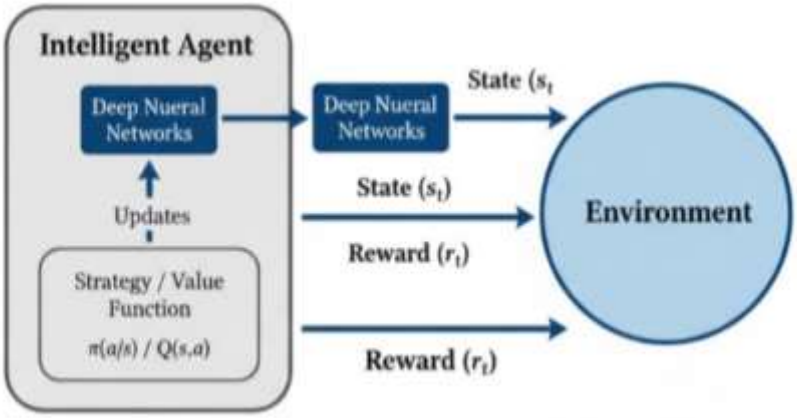

Figure 1 Deep Reinforcement Learning (DRL) Concept Framework.

2. The Evolution of Path Planning Algorithms: From Traditional to Intelligent

The development of AMR navigation has transitioned from deterministic geometric models to learning-based intelligent paradigms.

2.1 Traditional Algorithms and Hybrid Constraints

Before DRL, the field was dominated by deterministic models. While foundational, their inherent limitations necessitated a more adaptive approach. Table 1 breaks down these classic paradigms.

- *The Search-Based Guard: Dijkstra* (Dijkstra et al., 1976) *and A** (Hart et al., 1968) *are the gold standards for optimality. Yet, as the environment scales or becomes dynamic, their computational footprint grows exponentially.*
- *The Stochastic Alternative: RRT* (LaValle, 2006) *and PRM excel in high-dimensional voids but often yield "jittery" or suboptimal trajectories that are physically demanding for robot hardware.*
- *The Reactive Layer: APF* (Khatib, 1986) *allows for rapid reflex-like avoidance but is famously plagued by local minima, where a robot gets "stuck" simply because the math cancels out.*

ISSN: 3008-1394

Table 1 Comparison Of Traditional Path Planning Algorithms.

Algorithm Class	Typical Algorithms	Core Advantage	Primary Limitation
Search-Based	Dijkstra(Dijkstra et al., 1976), A*(Hart et al., 1968)	Guaranteed Optimality	Computational Gridlock
Sampling-Based	RRT (LaValle, 2006), PRM	Handles Complexity	Path Suboptimality
Reactive	APF(Khatib, 1986), DWA	Zero-Latency Response	Local Minima Trap

2.2 The DRL Shift: From Computation to Learning

The transition to DRL is essentially a move toward Representation Learning (Mnih et al., 2015). Rather than engineers deciding which features matter, the robot learns to interpret raw LiDAR spikes or pixel arrays directly (Kahn et al., 2021; Y. Zhu et al., 2025). This bypasses the information loss typical of manual abstraction.

3. Deep Reinforcement Learning Methodologies

In the field of DRL-based path planning, research primarily revolves around three major categories of algorithms, each with a different focus on handling value and policy functions.

3.1 Core Algorithmic Paradigms

The choice of DRL algorithm dictates how the agent learns. The main families are summarized in Table 2. The fundamental goal of DRL is to find a policy π that maximizes the expected cumulative reward G_t:.

$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right]$$

where γ ∈ [0,1] is the discount factor.

- *Value-Focus (DQN & Successors)* (Hasselt et al., 2015)*: By estimating the "worth" of discrete actions, DQN proved DRL's viability. However, its discrete nature often results in jerky, stop-and-go robot motion.*
- *Stability-Focus (PPO/TRPO)* (Schulman et al., 2017)*: PPO's clipped objective is the industry favorite for a reason: it prevents the robot from "forgetting" everything it learned due to a single bad update* (Xie et al., 2020)*.*
- *Efficiency-Focus (SAC/TD3)* (Zhang & Han, 2025)*: SAC is arguably the current state-of-the-art. By maximizing entropy, it encourages the robot to explore the environment more thoroughly, defined by the loss:*

$$L(\phi) = \mathbb{E}_{(s,a)\sim\mathcal{D}}[\frac{1}{2}(Q_{\phi}(s,a) - (r + \gamma\mathbb{E}_{s'}[V_{\bar{\phi}}(s')]))^2]$$

ISSN: 3008-1394

Table 1 Comparative Anylysis Of Key DRL Algorithms

Algorithm	Action Space	Sample Efficiency	Stability	Typical Use Case
DQN	Discrete	Low	Moderate	Simple grid-world navigation
PPO	Continuous	Moderate	High	Dynamic obstacle avoidance
TD3	Continuous	Moderate	Moderate	High-speed racing robots
SAC	Continuous	High	High	Complex multi-sensor fusion

3.2 Key Architectural Components

The performance of a DRL agent is heavily dependent on its neural network architecture, which processes sensory input and outputs actions. Fig. 2 illustrates a typical architecture.

- *Convolutional Neural Networks (CNNs) for Perception: For processing visual data from cameras or grid-based data from LiDAR, CNNs are the standard choice. Their convolutional layers are adept at extracting hierarchical spatial features, such as edges, corners, and object shapes, which are essential for identifying obstacles and navigable space.*
- *Recurrent Neural Networks (RNNs) for Temporal Reasoning: In dynamic environments, understanding the history of observations is crucial for predicting the movement of other agents. RNNs, particularly Long Short-Term Memory (LSTM) units, are designed to process sequential data. They can be integrated into the DRL agent's architecture to maintain a memory of past states, enabling it to reason about the velocity and intent of dynamic obstacles.*

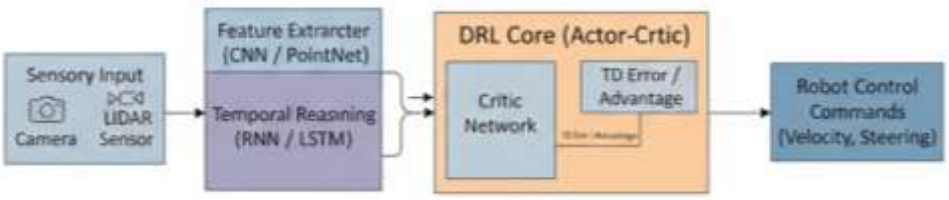

Figure 2 A Typical DRL Agent Architecture for Navigation.

3.3 Critical Training Strategies

Beyond the core algorithm, several strategies are crucial for successful training.

- *Reward Shaping: In many robotics tasks, the primary reward (e.g., reaching the goal) is sparse, making learning difficult. Reward shaping involves designing an auxiliary reward function that provides more frequent feedback. For instance, an agent can be rewarded for reducing its distance to the goal or penalized for getting too close to obstacles, guiding the learning process more effectively.*
- *Curriculum Learning: Instead of training the agent on the final, complex task from the beginning, curriculum learning involves starting with a simpler version of the task and gradually increasing the difficulty. For path planning, this could mean starting in an environment with few, static obstacles and progressively adding more obstacles and dynamic elements. This staged approach helps the agent to learn foundational skills first, leading to faster and more stable convergence on the final task.*

ISSN: 3008-1394

4. Confronting the "Reality Gap" and Future Frontiers

Despite significant progress, the application of DRL in mobile robot path planning still faces numerous challenges, which also represent important directions for future research.

4.1 Current Core Challenges

- *Robustness and Generalization: The adaptability of existing algorithms in complex, dynamic, and unknown environments is generally insufficient. Most research is still confined to specific or structured environments, lacking a universal avoidance method for high-density dynamic obstacles like crowds* (Da et al., 2025).
- *Data Efficiency and Training Complexity: DRL algorithms rely on massive amounts of high-quality training data, and real-world data collection is expensive and time-consuming. At the same time, the algorithms are highly sensitive to hyperparameters, making training convergence difficult and reproducibility poor* (Loquercio, 2023; Müller & Kudenko, 2025).
- *The Simulation-to-Reality (Sim2Real) Gap: Policies that perform excellently in simulation often see a sharp decline in performance when transferred to a physical robot, due to differences in dynamics, sensor noise, etc., between the physical world and the simulation* (Shakerimov et al., 2023).
- *Safety and Interpretability: For AMRs in human-centric spaces, safety is non-negotiable* (Gu et al., 2024). *Constrained MDPs (CMDP): Reward-based constraints* (Raji & Dobbe, 2023).*Control Barrier Functions (CBF): Integrating formal control theory with DRL for collision-free guarantees* (Dalal et al., 2018; Saunders et al., 2017).
- *Multi-Robot Collaboration: Existing research mostly remains at the level of distributed, decoupled planning. There is a lack of scalable, decentralized DRL frameworks that can accomplish complex collaborative tasks, not just conflict avoidance.*

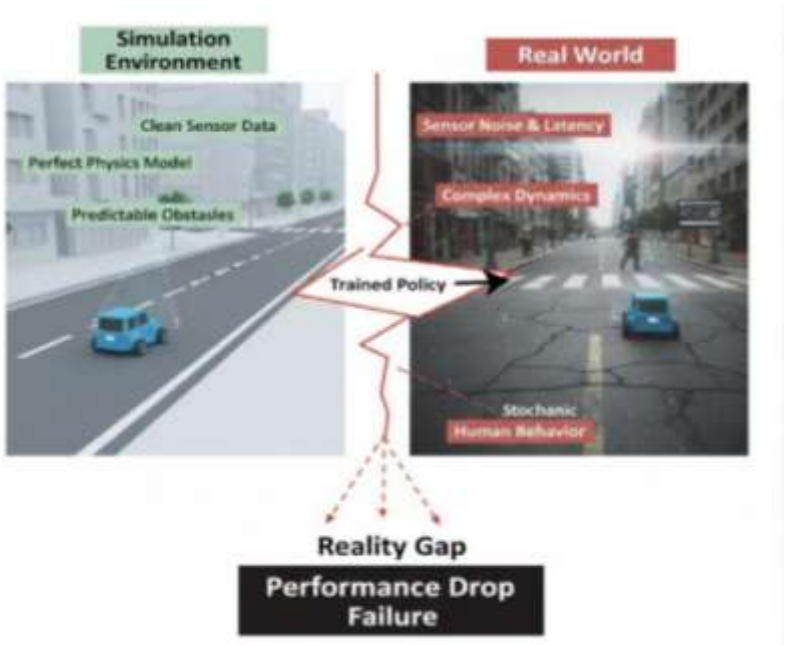

Figure 3 The Sim2Real Gap Challenge.

ISSN: 3008-1394

4.2 Future Research Directions

- *Foundation Models (LLMs/VLMs)* (Brohan et al., 2023)*: Using GPT-4V or RT-2 for semantic reasoning in navigation* (Huang et al., 2022; Shah et al., 2023).*Socially-Aware Navigation: Adhering to human social norms* (Everett et al., 2021; Okal & Arras, 2016).
- *Efficient Sim2Real Transfer Learning: Developing more powerful transfer learning, domain adaptation, and using generative AI (like GANs, diffusion models) to synthesize high-fidelity training data are key to bridging the gap between simulation and reality.*
- *Interpretable, Safe, and Ethical DRL: Integrating explainable AI (XAI), formal verification, and ethical constraints into DRL algorithm design is crucial to make the robot's decision-making process transparent, traceable, trustworthy, and compliant with societal expectations.*
- *Decentralized Multi-Robot Collaboration: Researching decentralized path planning frameworks under constraints such as limited communication, heterogeneity, and failures to achieve robust, scalable multi-robot collaboration*(Agal & Odedra, 2025; Yang, 2021).

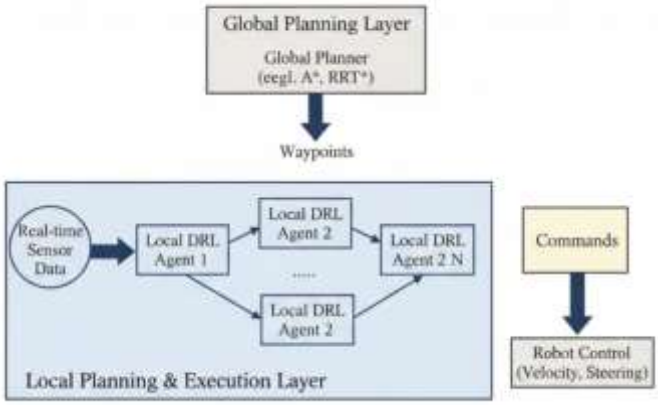

Figure 4 Hierarchical and Collaborative Path Planning Framework.

5. Conclusion

Reviewing 33 seminal works reveals that while the "perceptual" battle has been largely won, the "generalization" war continues. The path forward lies in hybrid models—systems that marry the raw power of DRL with the verifiable safety of classical control and the high-level reasoning of Foundation Models.

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.

ISSN: 3008-1394

Conflict of Interest

The authors have no conflicts of interest to declare.

References

[1] Agal, S., & Odedra, N. D. (2025). Decentralized reinforcement learning for scalable embodied intelligence in robotic swarms. *Emb Intell Robot*. https://www.researchgate.net/profile/Sanjay-Agal/publication/397090848_Decentralized_reinforcement_learning_for_scalable_embodied_intelligence_in_robotic_swarms/links/690ac9b2a2b691617b697b71/Decentralized-reinforcement-learning-for-scalable-embodied-intelligence-in-robotic-swarms.pdf

[2] Anwer, S., Hosen, M. S., Khan, D. S., Oluwabusayo, E., Folorunso, M., & Khan, H. (2024). Revolutionizing the global market: An inclusion of AI the game changer in international dynamics. *Migration Letters*, *21*(S13), 54–73.

[3] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Dabis, J., Finn, C., Gopalakrishnan, K., Hausman, K., Herzog, A., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jackson, T., Jesmonth, S., Joshi, N. J., Julian, R., Kalashnikov, D., Kuang, Y., … Zitkovich, B. (2023). *RT-1: Robotics transformer for real-world control at scale* (arXiv:2212.06817). arXiv. https://doi.org/10.48550/arXiv.2212.06817

[4] Da, L., Turnau, J., Kutralingam, T. P., Velasquez, A., Shakarian, P., & Wei, H. (2025). *A survey of sim-to-real methods in RL: Progress, prospects and challenges with foundation models* (arXiv:2502.13187). arXiv. https://doi.org/10.48550/arXiv.2502.13187

[5] Dalal, G., Dvijotham, K., Vecerik, M., Hester, T., Paduraru, C., & Tassa, Y. (2018). *Safe exploration in continuous action spaces* (arXiv:1801.08757). arXiv. https://doi.org/10.48550/arXiv.1801.08757

[6] Dijkstra, E. W., Dijkstra, E. W., Dijkstra, E. W., Informaticien, E.-U., & Dijkstra, E. W. (1976). *A discipline of programming* (Vol. 613924118). prentice-hall Englewood Cliffs. http://web.cecs.pdx.edu/~black/AdvancedProgramming/Lectures/Smalltalk%20II/Dijkstra%20on%20Hamming's%20Problem.pdf

[7] Everett, M., Chen, Y. F., & How, J. P. (2021). Collision avoidance in pedestrian-rich environments with deep reinforcement learning. *Ieee Access*, *9*, 10357–10377.

[8] Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., & Knoll, A. (2024). A review of safe reinforcement learning: Methods, theories, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 11216–11235.

[9] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). *Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor* (arXiv:1801.01290). arXiv. https://doi.org/10.48550/arXiv.1801.01290

[10] Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107.

ISSN: 3008-1394

[11] Hasselt, H. van, Guez, A., & Silver, D. (2015). *Deep reinforcement learning with double Q-learning* (arXiv:1509.06461). arXiv. https://doi.org/10.48550/arXiv.1509.06461

[12] Huang, W., Xia, F., Xiao, T., Chan, H., Liang, J., Florence, P., Zeng, A., Tompson, J., Mordatch, I., Chebotar, Y., Sermanet, P., Brown, N., Jackson, T., Luu, L., Levine, S., Hausman, K., & Ichter, B. (2022). *Inner monologue: Embodied reasoning through planning with language models* (arXiv:2207.05608). arXiv. https://doi.org/10.48550/arXiv.2207.05608

[13] Huber, J., Hélénon, F., Watrelot, H., Amar, F. B., & Doncieux, S. (2024). Domain randomization for sim2real transfer of automatically generated grasping datasets. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 4112–4118. https://ieeexplore.ieee.org/abstract/document/10610677/

[14] Kahn, G., Abbeel, P., & Levine, S. (2021). Badgr: An autonomous self-supervised learning-based navigation system. *IEEE Robotics and Automation Letters*, *6*(2), 1312–1319.

[15] 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

[16] LaValle, S. M. (2006). *Planning algorithms*. Cambridge university press. https://books.google.com/books?hl=en&lr=&id=-PwLBAAAQBAJ&oi=fnd&pg=PT7&dq=planning+algorithms+lavalle&ots=0jDv3rsolt&sig=6oa27J2DQGVFeSZyeDWZSlspIwE

[17] Lin, S., Liu, A., Wang, J., & Kong, X. (2022). A review of path-planning approaches for multiple mobile robots. *Machines*, *10*(9), 773.

[18] Loquercio, A. (2023). *Agile autonomy: Learning high-speed vision-based flight* (Vol. 153). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-27288-2

[19] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., & Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

[20] Müller, H., & Kudenko, D. (2025). *Improving the effectiveness of potential-based reward shaping in reinforcement learning* (arXiv:2502.01307). arXiv. https://doi.org/10.48550/arXiv.2502.01307

[21] Okal, B., & Arras, K. O. (2016). Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2889–2895. https://ieeexplore.ieee.org/abstract/document/7487452/

[22] Pitkevich, A., & Makarov, I. (2024). A survey on sim-to-real transfer methods for robotic manipulation. *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, 000259–000266. https://ieeexplore.ieee.org/abstract/document/10737545/

ISSN: 3008-1394

[23] Raji, I. D., & Dobbe, R. (2023). *Concrete problems in AI safety, revisited* (arXiv:2401.10899). arXiv. https://doi.org/10.48550/arXiv.2401.10899

[24] Saunders, W., Sastry, G., Stuhlmueller, A., & Evans, O. (2017). *Trial without error: Towards safe reinforcement learning via human intervention* (arXiv:1707.05173). arXiv. https://doi.org/10.48550/arXiv.1707.05173

[25] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms* (arXiv:1707.06347). arXiv. https://doi.org/10.48550/arXiv.1707.06347

[26] Shah, D., Osiński, B., & Levine, S. (2023). Lm-nav: Robotic navigation with large pre-trained models of language, vision, and action. *Conference on Robot Learning*, 492–504. https://proceedings.mlr.press/v205/shah23b

[27] Shakerimov, A., Alizadeh, T., & Varol, H. A. (2023). Efficient sim-to-real transfer in reinforcement learning through domain randomization and domain adaptation. *IEEE Access*, *11*, 136809–136824.

[28] Singh, R., Ren, J., & Lin, X. (2023). A review of deep reinforcement learning algorithms for mobile robot path planning. *Vehicles*, *5*(4), 1423–1451.

[29] Xie, Z., Clary, P., Dao, J., Morais, P., Hurst, J., & Panne, M. (2020). Learning locomotion skills for cassie: Iterative design and sim-to-real. *Conference on Robot Learning*, 317–329. http://proceedings.mlr.press/v100/xie20a.html

[30] Yang, X. (2021). Reinforcement learning for multi-robot system: A review. *2021 2nd International Conference on Computing and Data Science (CDS)*, 203–213. https://ieeexplore.ieee.org/abstract/document/9463292/

[31] Zhang, H., & Han, X. (2025). Off-policy asymptotic and adaptive maximum entropy deep reinforcement learning. *International Journal of Machine Learning and Cybernetics*, *16*(4), 2417–2429. https://doi.org/10.1007/s13042-024-02399-7

[32] Zhu, K., & Zhang, T. (2021). Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, *26*(5), 674–691.

[33] Zhu, Y., Wan Hasan, W. Z., Harun Ramli, H. R., Norsahperi, N. M. H., Mohd Kassim, M. S., & Yao, Y. (2025). Deep reinforcement learning of mobile robot navigation in dynamic environment: A review. *Sensors*, *25*(11), 3394. https://doi.org/10.3390/s25113394