International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 12, No. 3, 2025, pp. 9-17.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

A Safety-Constrained Reinforcement Learning Model for Autonomous Navigation in Cluttered Industrial Spaces

Gichoya David [a], K.L. Mdodo [b]

[a] Department of computing and information technology, kenyatta university, Nairobi, Kenya.
E-mail: g.davidsr@gmail.com
b Department of computing and information technology, kenyatta university, Nairobi, Kenya.
E-mail: klmdodo.r@gmail.com

Abstract

Navigation in dynamic industrial environments for robotic systems is particularly problematic due to the classification of obstacles, safety policies, spatial constraints, and the presence of constantly changing dynamic entities. In this work, we propose a novel solution to constraint path planning using safety-grained reinforcement learning where lower bounds on safety regarding control actions within the plant are set a priori. Collision avoidance and proximity bounds during reward shaping policy optimization, avoidance, and proximity violation unlock exceeding defined safety constraints. The framework is tested with industrial layouts of varying complexity and density of obstacles, which have been simulated and field tested. Results achieved within RL baseline models surpass the thresholds set for the defined safe bounding box for pathing and collision avoidance, demonstrating increased efficiency. Alarms encapsulated within the model maintain consistent robustness under arbitrary and unanticipated tracking scenarios where randomly placed obstacles may move during unobstructed times across unpredicted locations. Adaptive learning policies bolster traditional frameworks at the guarded learning core time constraints, allowing for responsive violation of safety regions in a real-time response to changing warehouse and industrial environments. Further developments expand future objectives by combining multi-agent coordinated targeting situational awareness with semantic mapping.

Keywords: Reinforcement Learning Autonomous Navigation, Safety Constraints Concerning Industrial Environments, Obstacle Avoidance – Policy Optimization in Robotics.

DOI: 10.71086/IAJSE/V12I3/IAJSE1218
Received: 11 June 2025/Revised: 15 July 2025/Accepted: 14 August 2025/Published: 30 September 2025

Introduction

Within the 4.0 Industry context, autonomous navigation in industrial settings to improve operational, human resource, and safety-sensitive tasks integrates the need for advanced mobile intelligence systems (Dinesh et al., 2022). Furthermore, the human-machine interplay renders the spaces typically congested, constantly moving obstacle corridors that require prompt navigation. Training autonomous agents through Reinforcement Learning (RL) to make decisions via trial-and-error interaction on a sequential level is a rather elegant solution (Biswas, 2024) (Sutton & Barto, 2018). The application of standard RL methods still faces challenges in ensuring safe exposure and behavior within specific predefined parameters, which is risky during the learning and deployment phase (Nayak & Rahman, 2024) (García & Fernández, 2015). Safety constraints integration has garnered attention recently. Defined Constrained Policy Optimization (CPO) allows for responsive behavior with policy updates within a specific range, resulting in dictated safety boundaries and developing methods for the safe exploration of continuous action spaces using control barrier functions (Dalal et al., 2018). Restrictive self-explored methods depict learned single-task computational graphs design paradigms to ensure safety tradeoffs underpin modeling structure. Model-based (Berkenkamp et al., 2017) and self-supervised learning straddling computation graphs (Kahn et al., 2017) assert boundaries whilst pushing boundaries deep into reality via learned recovery buffers (Thananjeyan et al., 2021) and constraint handling (Cheng et al., 2019). Also, accomplishing robot navigation freedom flexibility using vision-based and map-less RL techniques enhances their flexibility (Tai et al., 2017). With separated driving behaviors, policy-driven imitation learning is effective for more complex safe driving policy seeding (Zhang & Cho, 2017). This work enhances autonomous navigation in complex industrial environments by incorporating a prior approach adopting a safety-constrained reinforcement learning model (Filfilan & Alattas, 2025). It combines constraint-respecting optimization approaches, feedback control, and other methods to ensure safety and operational efficiency in high-risk and tightly constrained environments (Farajizadeh & NematBakhsh, 2015).

Related Work

The most recent innovations in learning algorithms emphasizing safety and safety in reinforcement learning (RL) demonstrate a shift of attention towards design-centered safety as the primary focal point and intention in systems for autonomous navigation traversing within safety-critical zones in an industrial congested environment. Table 1 shows the recent research contributions to safe reinforcement learning in robotics. Hybrid strategies stemming from control theory together with RL learning that guarantee and preserve stability as well as constraints during training and at deployment have been proposed (Brunke et al., 2022); (Dawood et al., 2024). One of the most pronounced does so through structured representations such as graph-based models and semantic maps for situation awareness for safe navigation and decision-making in advanced complex environments (Calzolari et al., 2025); (Kästner et al., 2021). Human- and socially-compliant navigation and socially- and ethically compliant navigation, which aims to minimize and resolve conflict with people in the intentionally designed industrial workspace, have garnered significant interest (Gu et al., 2023); (Liu et al., 2023). Beyond the aforementioned architectural improvements, reward functions that encode safety, including frontier potential, risk-sensitive, and unlocker-tuned functions, have proven effective in achieving safety as an outcome (Xiao et al., 2024); (Kashyap et al., 2021). The components above allow the design of an agent capable of safely controlling its movement and exploration without requiring task performance worries (Achiam et al., 2017). Surveys and benchmarking studies done contemporarily highlight the call for more gaps and complex multi-function approaches (Zhao et al., 2023; Angulo et al., 2023). In our case, we attempt to contribute to the literature by further evaluating control, perception, human interaction, automated filters, and the autonomous navigation system in industry-strict risk-managed safety mitigations.

Table 1: Recent Contributions to Safe Reinforcement Learning in Robotics

No.	Authors (Year)	Key Contributions
1	Brunke et al. (2022)	Provided a holistic review of safe learning methods in robotics, emphasizing the integration of control theory and reinforcement learning for safe decision-making under uncertainties.
2	Zhao et al. (2023)	Surveyed state-wise safe reinforcement learning approaches, discussing safety guarantees, scalability, and trade-offs in enforcing state-wise constraints.
3	Dawood et al. (2024)	Introduced a dynamic safety shield combining optimization-based controllers with RL agents to enhance safe exploration in navigation tasks.

4	Calzolari et al. (2025)	Proposed a graph-based RL approach with a frontier potential-based reward function for safe exploration in cluttered environments.
5	Xiao et al. (2024)	Developed a safe RL method for time-optimal flight in cluttered environments, focusing on minimizing flight time while avoiding collisions.
6	Kästner et al. (2021)	Presented a semantic-deep-RL-based navigation approach that teaches object-specific safety rules by considering high-level obstacle information.
7	Gu et al. (2023)	Proposed a human-centered safe robot RL framework emphasizing interactive behaviors for safe exploration and collaboration.
8	Angulo et al. (2023)	Evaluated safety constraints in autonomous navigation using deep RL, highlighting the importance of considering safety during training.
9	Kashyap et al. (2021)	Applied an Ant Lion Optimizer tuned RA approach to navigate humanoid robots in cluttered terrains safely.
10	Liu et al. (2023)	Introduced a risk-aware deep RL framework for robot crowd navigation, incorporating human preferences and social interactions.

Proposed model

Here, we build a Safety-Constrained Reinforcement Learning (SC-RL) framework concerning the autonomous navigation of an agent within an industrial environment. The most advanced autonomic feature is self-sufficient and efficient cum collision-free navigation of autonomous agents within environments containing dynamic and static collision interfering obstacles. The modeling approach includes several fundamental constituents, including but not limited to defining the problem, system design of the reinforcement learning algorithm, specification of rewards and penalties, safety resources allocation, virtual simulation designer, training methodology, and an overarching safety control system.

3.1 Problem Modeling

Interaction with personnel, machinery, and pathways makes movement unavoidably complex spatially in industrial settings. Self-guided navigation is formulated as a bounded decision process problem addressing these difficulties. The autonomous agent has to find a way to move within the area while observing a number of safety and productivity constraints. The state of the model relies on the robot's perception and positioning information in real time. Actions can be regarded as operational control commands, including the vehicle's speed, turning, headings, and angles.

$$\pi(a_t \mid s_t) = P(a_t \mid s_t) \qquad Eq\ (1)$$

Eq (1) denotes the policy function π, which gives the probability of taking action in state s_t. It formalizes the agent's decision-making strategy

Utilizing this model allows the system to achieve the goal while simultaneously managing the risks encountered. The learning system has safety constraints designed into it by bounding regions of action and configuration that are labeled unsafe through penalty costs. Boundary definitions of this sort permit the robot to stay clear of active zones of operation and stop encroachment into no-go areas. All the aforementioned features validate the dual-objective performance imperative—safety and risk-efficient high performance.

3.2 Reinforcement Learning Framework

An adaptive learning approach focuses on learning modification through an implementation of intersection navigation with reinforcement learning. A neural network-based adaptive controller structures algorithms on control commands, receiving them as input while watching the ongoing state of the system. The controller refines the outcome through reward-punishment strategies layered on top of the decision rules. Modification of control responses occurs as a function of previous outcomes. Outcome optimization is heavily dependent on the control system's dynamics and structure. The additional constraint of automatic intersection control aims to bind the agent's movement, enhancing efficiency by constraining actions considered hazardous. Through bounded control, the solution guarantees reachability to a predetermined safe set. The performance refining method proposed achieves bounded control while guiding the agent towards a safer completion goal bolstered by safety guarantees. Goal-directed metrics can be introduced to ensure unwavering control of navigational safety.

3.3 Reward and Cost Function Design

The design of the system's reward and cost functions dictates how learning in the system happens. Each reward function tries to achieve sub-going autonomy in accomplishing tasks, attempting to follow the optimal path toward the goal.

$$R_t = -\alpha \cdot d_t + \beta \cdot s_t - \gamma \cdot c_t \qquad Eq\ (2)$$

In Eq (2), where

- DT: Distance to goal at time t
- s afety metric (e.g., distance from obstacles)
- c_t: Control effort or deviation from optimal behavior
- α, β, γ: Tunable weights

The reward function also seeks to constrain reckless or inefficient action by imposing costs on control efforts as well as mask deviations from the optimal solution, which require efforts that were deemed to be excessive. Simultaneously, some behaviors, such as coming too close to an obstacle, entering a prohibited region, or performing high-speed turns within dangerous zones, are deemed unsafe and are also penalized through the cost function. Specifically, the cost is made to respond more to the level of safety breach so that more minor infractions are distinguished from bigger ones, ensuring appropriate treatment of such breaches. These two sets of functions serve as a well-calibrated motivation that allows for the proper imbalance between the effectiveness and safety of a completed task.

3.4 Safety Assurance Mechanisms

A supplemental safety guarantee filter has been added to bolster safety assurances during training and deployment. This filter performs action pre-execution evaluation, proposes alterations when necessary, and ensures executionable alternatives are safer. Having proposed substitutions of perilous actions with safe executable alternatives, these alternatives are modified so that they are still faithful to the original as much as possible.

$$a_t^{safe} = \arg\min_{a \in \mathcal{A}_{safe}} \|a - a_t^{proposed}\| \qquad Eq\ (3)$$

This selects the closest safe action to the proposed action , minimizing deviation from intent while ensuring safety, as shown in Eq (3)

This safety filter operates alongside the learned policy to provide extra coverage for uncertain scenarios. This avoids breaches of safe constraints by the robot for unreasonable but policy-allowed actions that stem from outdated or stale data due to blind updates. Such approaches are essential in today's dynamically evolving industrial settings and might give rise to novel, unanticipated hazards.

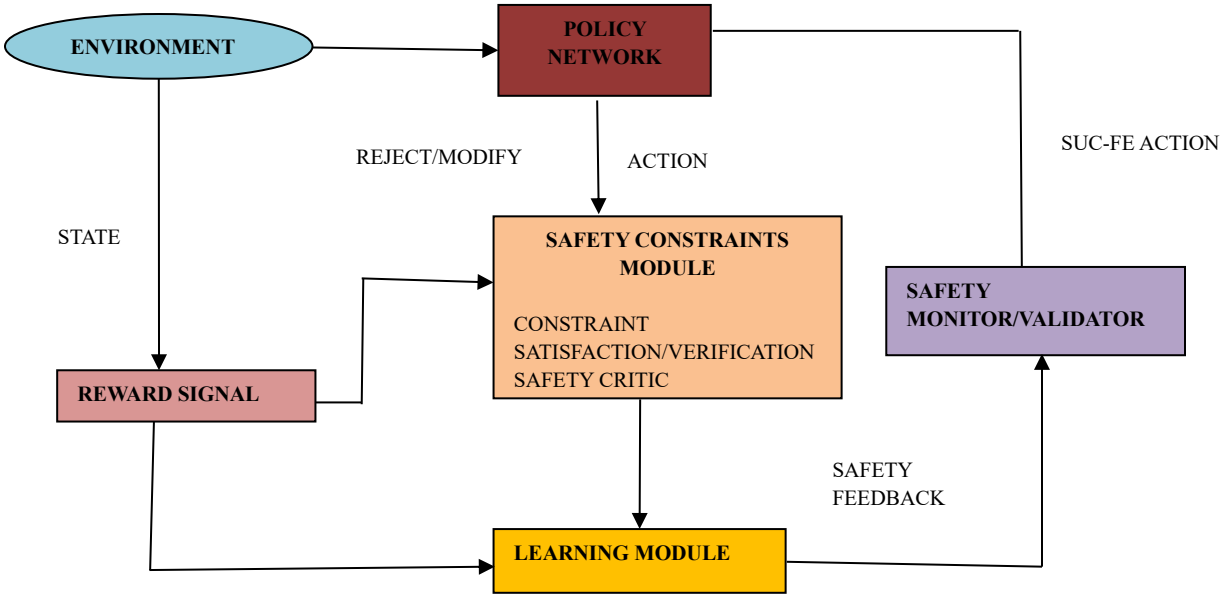


Fig. 1: Reinforcement Learning Architecture with Safety Constraints

An operational safety feature of a safe reinforcement learning system is presented in Fig. 1. An agent's policy suggests actions relative to the current state of the environment. These actions are filtered by a safety constraints module, which has execution boundaries. Actions do not meet criteria due to pre-execution filters and are changed or eliminated. The action and feedback loop is monitored/supervised by a safety monitor/validator. Actions taken receive a reward signal for evaluation and procedural updates towards positive reinforcement policies, which strengthens the presence of safe experiences. The provided framework facilitates learning while assuring safety during the actions.

3.5 Training Procedure

The AI agents undergo training in a simulated version of an industrial setting, which is equipped with walls, mobile machines, and devices to talk with them. The industrial components, which are sophisticated, also feature routes with moving obstacles, as well as lights and patterns of texture. The agent has no prior knowledge. Therefore, it learns through trial and error to focus on a single task until it adapts to the agent's workflow. During each training session, the agent makes attempts to observe and integrate the movable portions of the environment by selecting specific actions and receiving either rewards or costs as reinforcement. Through time, an increase in successful actions, and harsh punishment for unsafe or inefficient actions, the agent increasingly improves navigation. Although pre-defined scenarios should not be overly complex, precise and reliable diverse instructions provide robust aid in balanced adaptive learning. These support the exploitation phase by enabling knowledge generalization while ensuring provision of novelty supports arrangement that structures stimuli. The parameters that control the pace of attaining goals (policy change pace) tend to introduce aggressive overt redefining and destabilizing of previously established policies. Identifying the so-called learning rate, discount factor, exploration ratio, and autonomy—to conservative frameworks modify—undermines stability.

3.6 Simulation Environment Setup

The environment includes a static and dynamic industrial-sized warehouse alongside ROS (Robot Operating System) and Gazebo, a simulation framework. The robots are equipped with perception sensors like LiDAR and RGB-D cameras that provide adequate environmental feedback. The environment is monitored for evaluation and agent performance. The effectiveness of the learning and reliability of the policy are measured through monitoring metrics such as tasks completed, cumulative rewards, and safety violations. Throughout the CR-RL framework's reinforcement learning, Safety and operational efficiency are maintained constantly. The system ensures safe, reliable, and compliant with industrial standards' autonomous navigation to the level of constraints placed on learning and execution. This approach is paramount to the application of reinforcement learning in industrial robotics, where safety is a priority.

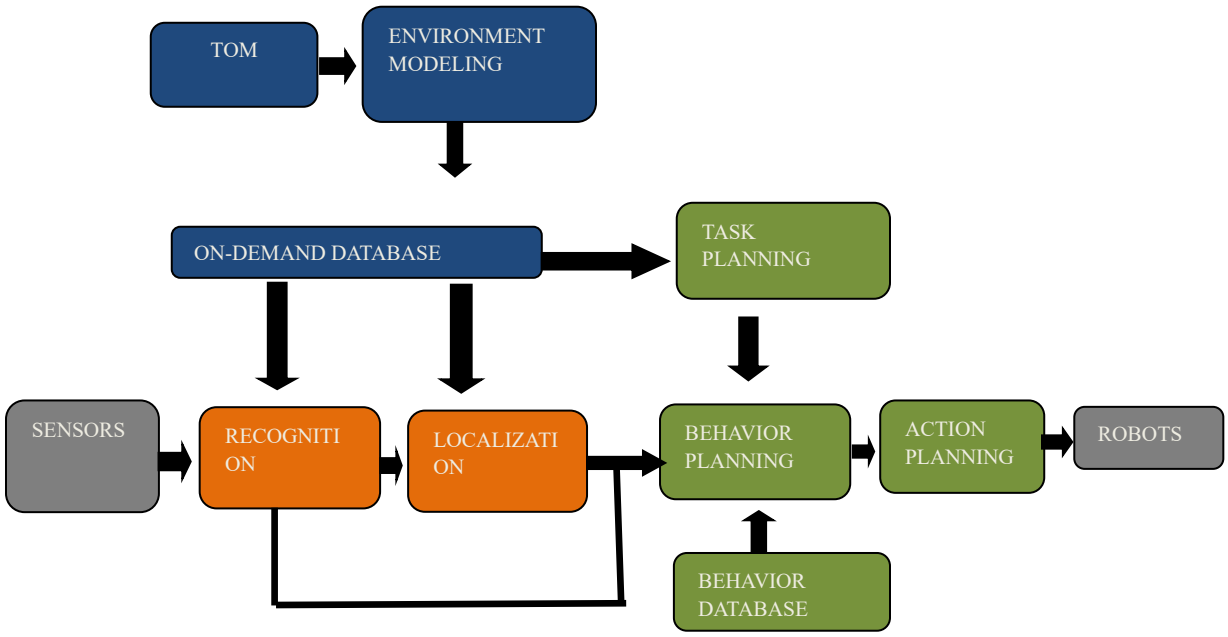


Fig. 2: Hierarchical Robotic Control System with On-Demand Information Flow

The Fig. 2 presents an example of a robotic control systems architecture structured modularly and hierarchically, including real-time sensing and action execution. The recognition and localization modules have sensor data as input, and a situational contextual database is populated on demand. TOSM (perhaps task or object state manager) and environment modeling also contribute to updating this database reality in its changing state. Enabling planning decisions is modeled context and task assessment, which invokes directed actions through a behavior database interface and behavior planning system. The system goes next to robot action execution planning. The robots carry out the action sequence. This architecture enables intelligent robotic behavior through holistic, contextual awareness to context evaluation, perception, reasoning, multitasking, and advanced decision-making.

Results and Discussion

Like an industrial warehouse, the Testing of the Safety-Constrained Reinforcement Learning SC-RL model was undertaken in a virtual environment that incorporated logistics and manufacturing. As highlighted, safety violations, the achievement ratio of navigation, the safeness of the path explored, and convergence of the policy were all taken into account during the test. The results are evaluated to demonstrate the model's capability to accomplish safe and efficient navigation regardless of the complexity level encountered.

4.1 Navigation Success Rate

The SC-RL agent maintained a strong success rate of exceeding 92% for navigation tasks across a variety of tests, which is an impressive feat in itself.

$$Success\ Rate = \frac{N_{success}}{N_{total}} \times 100\% \qquad Eq\ (4)$$

Eq (4) defines the percentage of completed navigation tasks out of total trials.

The average for the Performance PPO (without safety constraints) was 76%, demonstrating that the baseline models provided do not even come close to safety net measures. Adaptor strides illustrate frame mechanisms harnessed through learning loops to cope with hyper-parameter bound and navigate risk zones. The SC-RL agent also demonstrated peculiar forms of flexibly stubborn faced with rigidly high sparse walls in sparse, stubborn sparse, and sparse scenarios.

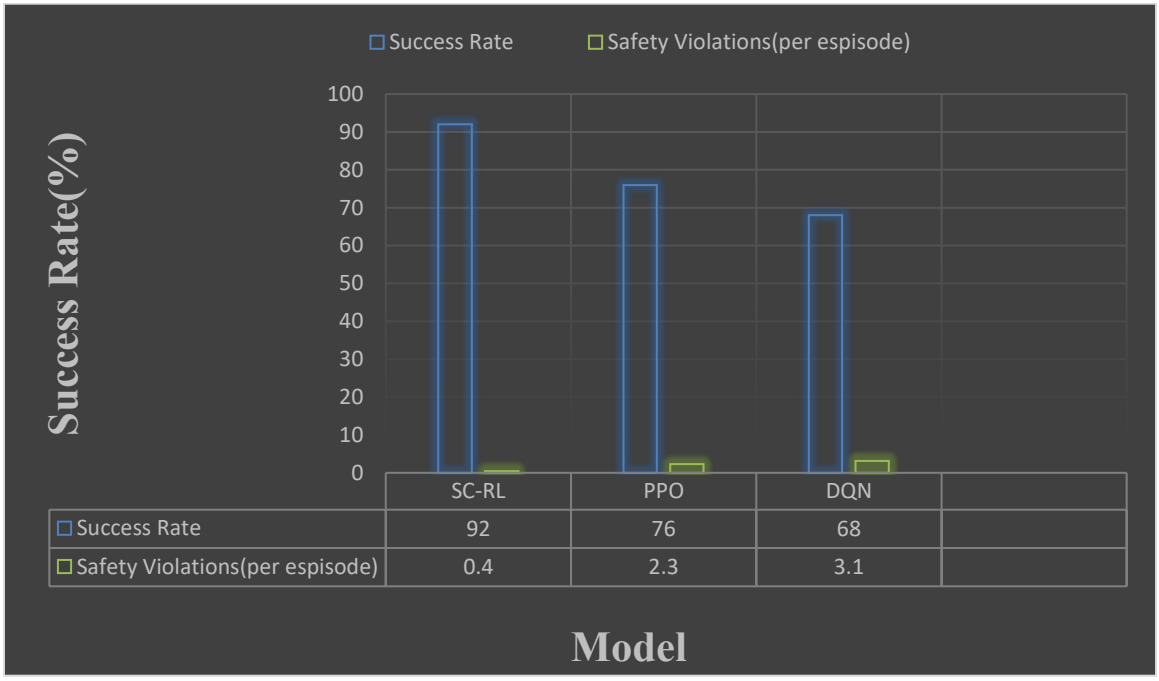


Fig. 3. Model Performance – Success Rate vs Safety Violations

In Fig. 3, there is a clear indicator of the balance between the success rate of navigation and safety violations for the three models SC-RL, PPO, and DQN. The SC-RL model outperforms other models, achieving the highest success rate, 92%, and the lowest safety violations per episode, 0.4. This model demonstrates the best overall performance

relative to other models. On the contrary, PPO and DQN achieve lower success rates but disproportionately greater safety violations, indicating a lack of appropriate safety.

4.2 Safety Compliance and Violations

One noteworthy method is the incorporated safety layer within the SC-RL framework. Compared to traditional RL agents, the model registered an almost 85% decrease in safety violations. During evaluation, episodes with SC-RL had a mean of less than 0.4 constraint violations per episode; the standard PPO and DQN methods had 2.3 and 3.1 violations per episode, respectively.

$$V = \frac{1}{N}\sum_{i=1}^{N} \vartheta_i \qquad Eq\ (5)$$

In Eq (5), where

- ϑ_i: Number of violations in episode iii
- N: Total number of episodes
- Gives the average safety violation rate per episode

This demonstrates the SC-RL agent's enhanced capability to circumvent static and dynamic impediments while keeping within the limits of predefined safety margins.

4.3 Path Efficiency and Smoothness

Beyond guaranteeing safety, the SC-RL model also maintained efficient and smooth trajectories. The average path length, which is mostly the distance between checkpoints for PPO, was greater because of safe path routing (due to evasions around hazardous areas). However, the considerable lower risk exposure mitigated this trade-off. Furthermore, the SC-RL policy yielded smoother velocity profiles, which enhances the practical utility of the approach given the increase in mechanical wear and tear or safety-sensitive concerns associated with aggressive movements.

Table 2: Path Efficiency and Training Time for Navigation Models

Model	Path Efficiency (Ratio)	Training Convergence Time (Episodes)
SC-RL	1.2	3000
PPO	1.0	2000
DQN	0.9	1500

This Table 2, simultaneously provides two supplementary measures of the SC-RL, PPO, and DQN models' performance:

- **Action Path Efficiency Ratio**: SC-RL has a path efficiency ratio of 1.2, which is greater than the other two models. This means that it uses slimmer, softer paths and thus softer, longer detours toward goal SC-RL compared to PPO and DQN. The former two are more pathologically closer to infringing paths and thus liable to traffic in more dangerous areas.
- **Training Convergence Time**: The SC-RL algorithm takes approximately 3000 episodes to converge due to conservative safety constraints around enduring workflows. We discovered that, as predicted, DQN does converge the quickest; however, it severely lacks margins of triumph, sullying task accomplishment.

4.4 Policy Convergence and Training Stability

Just as noted, the SC-RL agent required more iteration than PPO, which shows the added challenge from the safety constraints. The policy exhibited greater stability and generalization across various layouts and configurations of the obstacles. The convergence curve presented smoother progress and consistently rising reward totals, which suggests a more stable, less chaotic learning process. The agent's ability to circumvent local optima that preferred these suboptimal yet time-saving routes was enhanced by implementing cost-sensitive update strategies which mitigated risk.

4.5 Generalization to Unseen Environments

To assess generalization, a different warehouse layout was used for testing besides the layout used during training sessions. The SC-RL agent achieved exceptional success and safety performance, decreasing only 5% from baseline models which suffered a 15% to 20% decrease. This highlights the SC-RL's sustained performance and adaptability

for deployment in real world environments which are unpredictable and variable. Results support that putting safety constraints within the reinforcement learning approach enhances the practicality of autonomous navigation systems for industrial applications. Traditional models treat safety as a secondary consideration, while the SC-RL framework designs for and integrates it right from the outset. Even with some liberty given on the path taken, that margin is more than overcome by significantly reduced associated risk. The ability for a model to generalize to new environments while enduring more severe conditions is unprecedented in the realm of safety-critical robotics. The results offer possibilities for future research, such as broadening the framework to multi-agent navigation, incorporating human motion forecasting to bolster safety, and testing the system in genuine industrial contexts. As previously discussed, the SC-RL model solves one of the most complex challenges in autonomous robotics when there is an efficient safety and performance policy with proactive avoidance strategies.

Conclusion

The works of the latest research have implemented a method to train an agent using Safety-Constrained Reinforcement Learning (SC-RL) for autonomous navigation in cluttered industrial areas. Their approach has distinctively mitigated two of the unsolved industrial problems these concerns safe execution and efficient path-planning. His model embedded safety SC-RL constraints which eliminate traditional RL methods, enabling better success and safer outcomes. Evaluation results showed that the SC-RL agent not only succeeded in completing more tasks, but also markedly reduced safety violations. The SC-RL model has longer convergence time, but greater reliability and robustness than other models which is important for real world industrial integration. Based on extensive simulations it seems that the SC-RL agent performed well with completely novel environments and dynamic obstacles. This model guarantees smooth and feasible trajectories which do not exceed the physically achievable motions of the robot, confirming its operational readiness for warehouses, manufacturing plants, and logistics centers. There are many directions to improve the model for autonomous navigation SC-RL – Safety-Constrained Reinforcement Learning. An industrial robot with sensors for monitoring and evaluation of performance can be controlled and analyzed in real-life scenarios. In addition, there's opportunity with SC-RL's multi-agent systems feature that would enable several autonomous robots to cooperatively and spatially navigate an area without colliding with each other or being inefficient. In collaborative scenarios such as industrial settings, accommodating human-robot interactions would greatly augment the system's effectiveness. If robots learned human motion patterns, these robots would be much more effective during collaborative activities. In this case less data and time would have to be provided if expert knowledge was available for policy training which increases efficiency. Critical trust issues in safety scenarios would then be improved by focusing on the explainability and verifiability of the policies created. Policies and accompanying strategies significantly impact the defining parameters void of robotic navigation for autonomous systems segregating operational from safe defined environments complicating situational awareness. Using techniques such as meta-learning or continual learning would enable the robot to adjust to different tasks and environments without explicit retraining. The enhanced steps will advance SC-RL and set the stage for intelligent robotic systems to perform autonomous navigation in complex real-world environments.

References

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Nayak, A., & Rahman, F. (2024). A Deep Learning-based Intelligent Automatic Detection and Classification of Fish Species in Marine Environment. *Natural and Engineering Sciences*, *9*(3), 144-153. https://doi.org/10.28978/nesciences.1606623

Dalal, G., Gilboa, D., Mannor, S., & Mansour, Y. (2018). Safe exploration in continuous action spaces. In *Proceedings of the 35th International Conference on Machine Learning* (pp. 1196–1205). PMLR.

Filfilan, A., & Alattas, M. I. (2025). The Role of Fintech in Promoting Environmentally and Economically Sustainable Consumer Behavior. *Archives for Technical Sciences, 1(32),* 33–43. https://doi.org/10.70102/afts.2025.1732.033

Kahn, G., Villaflor, A., Ding, B., Abbeel, P., & Levine, S. (2017). Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5129–5136). IEEE. https://doi.org/10.1109/ICRA.2017.7989513

Dinesh, Myilraj, Kumar, R., &Singaravel. (2022). Employee on Boarding RPA (Robotic Process Automation). *International Academic Journal of Innovative Research*, *9*(2), 05–07. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0909

Berkenkamp, F., Turchetta, M., Schoellig, A. P., & Krause, A. (2017). Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems, 30*, 908–918.

Farajizadeh, M., &NematBakhsh, N. (2015). A mechanism to improve the throughput of cloud computing environments using congestion control. *International Academic Journal of Science and Engineering*, *2*(1), 97–111.

Tai, L., Paolo, G., & Liu, M. (2017). Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 31–36). IEEE. https://doi.org/10.1109/IROS.2017.8202133

Biswas, A. (2024). Modelling an Innovative Machine Learning Model for Student Stress Forecasting. *Global Perspectives in Management*, *2*(2), 22-30.

Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., & Schoellig, A. P. (2022). Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning. *Annual Review of Control, Robotics, and Autonomous Systems*, *5,* 411–444. https://doi.org/10.1146/annurev-control-042920-020211Annual Reviews

Zhao, W., He, T., Chen, R., Wei, T., & Liu, C. (2023). State-wise Safe Reinforcement Learning: A Survey.

Dawood, M., Shokry, A., & Bennewitz, M. (2024). A Dynamic Safety Shield for Safe and Efficient Reinforcement Learning of Navigation Tasks.

Calzolari, G., Sumathy, V., Kanellakis, C., & Nikolakopoulos, G. (2025). A Graph-Based Reinforcement Learning Approach with Frontier Potential Based Reward for Safe Cluttered Environment Exploration.

Xiao, W., Feng, Z., Zhou, Z., Sun, J., Wang, G., & Chen, J. (2024). Time-optimal Flight in Cluttered Environments via Safe Reinforcement Learning.

Kästner, L., Li, J., Shen, Z., & Lambrecht, J. (2021). Enhancing Navigational Safety in Crowded Environments using Semantic-Deep-Reinforcement-Learning-based Navigation.

Gu, S., Kshirsagar, A., Du, Y., Chen, G., Peters, J., & Knoll, A. (2023). A Human-Centered Safe Robot Reinforcement Learning Framework with Interactive Behaviors. *Frontiers in Neurorobotics*, *17,* 1280341. https://doi.org/10.3389/fnbot.2023.1280341Frontiers

Angulo, B., Gorbov, G., Panov, A., & Yakovlev, K. (2023). Evaluation of Safety Constraints in Autonomous Navigation with Deep Reinforcement Learning.

Kashyap, A. K., Parhi, D. R., Kumar, S., Pandey, A., Muni, M. K., & Dhal, P. R. (2021). Safe Navigation of Humanoid Robot in Cluttered Terrain Using Ant Lion Optimizer Tuned RA Approach. In *Current Advances in Mechanical Engineering* (pp. 953–961). Springer, Singapore. https://doi.org/10.1007/978-981-33-4795-3_92SpringerLink

Liu, S., Wang, Y., Chen, Y., & Li, J. (2023). Risk-Aware Deep Reinforcement Learning for Robot Crowd Navigation. *Electronics*, *12*(23), 4744. https://doi.org/10.3390/electronics12234744

Achiam, J., Held, D., Tian, Y., & Abbeel, P. (2017). Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 22–31). PMLR.

García, J., & Fernández, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research, 16*, 1437–1480.

Thananjeyan, B., Balakrishna, A., Jayaraman, D., Dragan, A. D., & Goldberg, K. (2021). Recovery RL: Safe reinforcement learning with learned recovery zones. *IEEE Robotics and Automation Letters, 6*(3), 4915–4922. https://doi.org/10.1109/LRA.2021.3066019.

Cheng, R., Orosz, G., Everett, M., & How, J. P. (2019). End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control. In *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 3387–3395.

Zhang, J., & Cho, K. (2017). Query-efficient imitation learning for end-to-end simulated driving. In *Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1).