World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:17, No:12, 2023

A Control Model for Improving Safety and Efficiency of Navigation System Based on Reinforcement Learning

Almutasim Billa A. Alanazi, Hal S. Tharp

***Abstract*—**Artificial Intelligence (AI), specifically Reinforcement Learning (RL), has proven helpful in many control path planning technologies by maximizing and enhancing their performance, such as navigation systems. Since it learns from experience by interacting with the environment to determine the optimal policy, the optimal policy takes the best action in a particular state, accounting for the long-term rewards. Most navigation systems focus primarily on "arriving faster," overlooking safety and efficiency while estimating the optimum path, as safety and efficiency are essential factors when planning for a long-distance journey. This paper represents an RL control model that proposes a control mechanism for improving navigation systems. Also, the model could be applied to other control path planning applications because it is adjustable and can accept different properties and parameters. However, the navigation system application has been taken as a case and evaluation study for the proposed model. The model utilized a Q-learning algorithm for training and updating the policy. It allows the agent to analyze the quality of an action made in the environment to maximize rewards. The model gives the ability to update rewards regularly based on safety and efficiency assessments, allowing the policy to consider the desired safety and efficiency benefits while making decisions, which improves the quality of the decisions taken for path planning compared to the conventional RL approaches.

***Keywords*—**Artificial intelligence, control system, navigation systems, reinforcement learning.

I. Introduction

NAVIGATION systems have been used for a long time, beginning with the magnetic compass, and progressing to the present with global positioning system technology (GPS). However, most navigation systems prioritize “arriving faster” [1], [2] while ignoring important safety and efficiency considerations. Arriving in a short time interval is a complex and vital subject for any navigation system. Yet, keeping safety and efficiency in mind is critical.

Some drivers have limited abilities, such as an impaired vision, and others require hospital access when traveling from one location to another, particularly on long-distance trips. For example, some older drivers dread driving in adverse weather conditions such as rain because it poses a significant safety risk to them. According to the Federal Highway Administration - Department of Transportation in the United States (U.S.), every year, around 5,891,000 automobiles are involved in accidents. Weather accounts for around 21% of these crashes (almost 1,235,000). Weather-related accidents can occur in weather conditions such as rain, sleet, snow, fog, strong crosswinds, or blowing snow/sand/debris. Approximately 5,000 individuals are killed, and over 418,000 are injured in weather-related crashes each year [3].

In terms of efficiency, especially for Electrical Vehicle (EV) drivers. As mentioned early, the navigation systems focus on arriving faster, which does not explicitly mean short distances or less energy consumption. Therefore, EV drivers might face some charging challenges for long distance travel. Even though EVs are becoming increasingly popular, as in 2021, the global sales of electric vehicles reached 6.6 million, accounting for about 9% of the global car market [4]. Also, according to the International Energy Agency (IEA), EV sales are expanding exponentially, exceeding 10 million by 2022. EV sales have tripled in three years, reaching roughly 4% in 2020 to 14% in 2022 of the global sales (see Fig. 1) [5].

So, EV drivers need to consider some features while planning their trips, like charging stations and avoiding roads that consume more energy, such as congested traffic or mountain driving, which also applies to gas vehicles. However, EV charging is more challenging compared to a gas fill-up, since EV charging takes more time. For instance, DC fast charging technology is currently the fastest. It requires a 480-volt connection, making DC charging impractical to use in homes, and it is not available on all-electric car models; it contributes up to 10 miles of range every minute of charging time, depending on battery type, charger arrangement, and circuit capacity. That is about 40 minutes for a 400-mile trip [6]. Unfortunately, the present navigation systems primarily focus on arriving faster (shortest time). The drivers could use some manual steps to make sure their needs are met, for example, searching for nearby EV stations along the way and then adding it as "Add a Stop" or using Apps separately from the navigations system like ChargePoint, Evgo, or PlugShare to adjust their trip plans based on the EV station locations.

RL mimics how humans and animals naturally learn the optimal behavior in an environment to obtain the maximum rewards. As an example from nature that illustrates accounting for safety and efficiency, a female European honey buzzard bird equipped with a satellite tracking system made a journey from

Almutasim Billa A. Alanazi is a Ph.D. student with the Electrical and Computer Engineering Department, The University of Arizona, Tucson, AZ USA (e-mail: abalanazi@arizona.edu).

Hal S. Tharp is Associate Department Head of Electrical and Computer Engineering and Associate Professor of Electrical and Computer Engineering with the Electrical and Computer Engineering Department, The University of Arizona, Tucson, AZ USA (e-mail: tharp@arizona.edu).

South Africa to Finland. The bird traveled almost 10,000 kilometers at a 230 km/day speed in about 42 days [7]. Interestingly, the bird took safety and efficiency under consideration, as it avoided the Mediterranean Sea and the desert along the way (unsafe regions) and followed the Nile River in case of getting thirsty (fuel efficiency). Fig. 2 shows the bird's map journey.

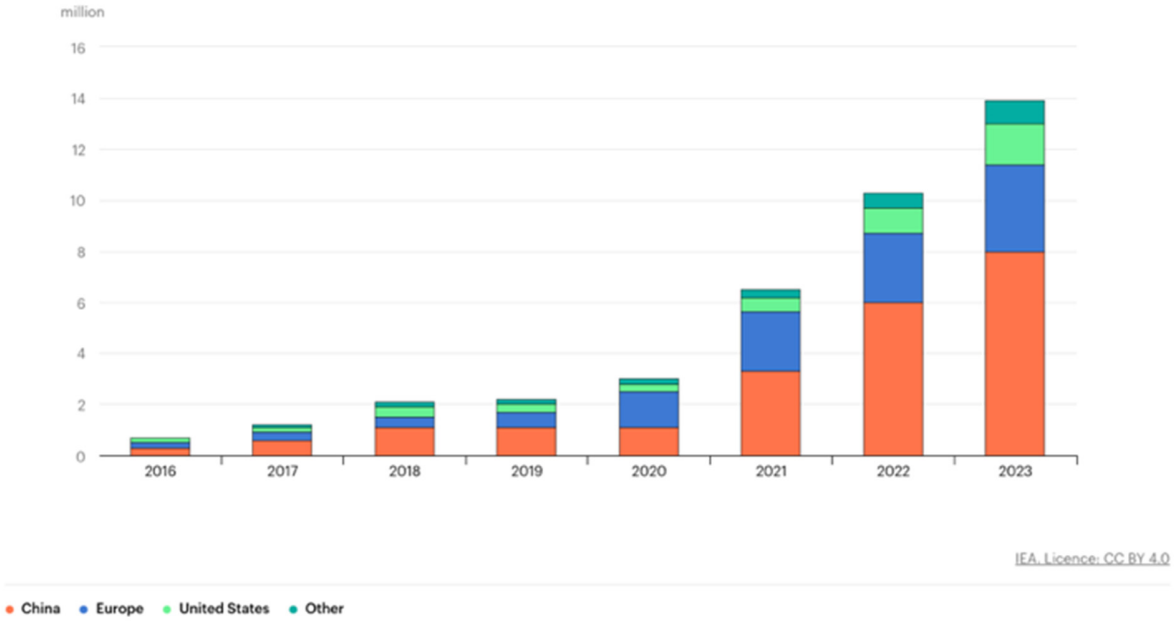


Fig. 1 Electrical car sales, 2016 – 2023 [5]

Fig. 2 Bird's path journey

Consistent with the previous discussions, this paper presents an RL model with more control ability to improve the navigation system and take more features like safety and efficiency under consideration while generating the navigation path.

II. Literature Review

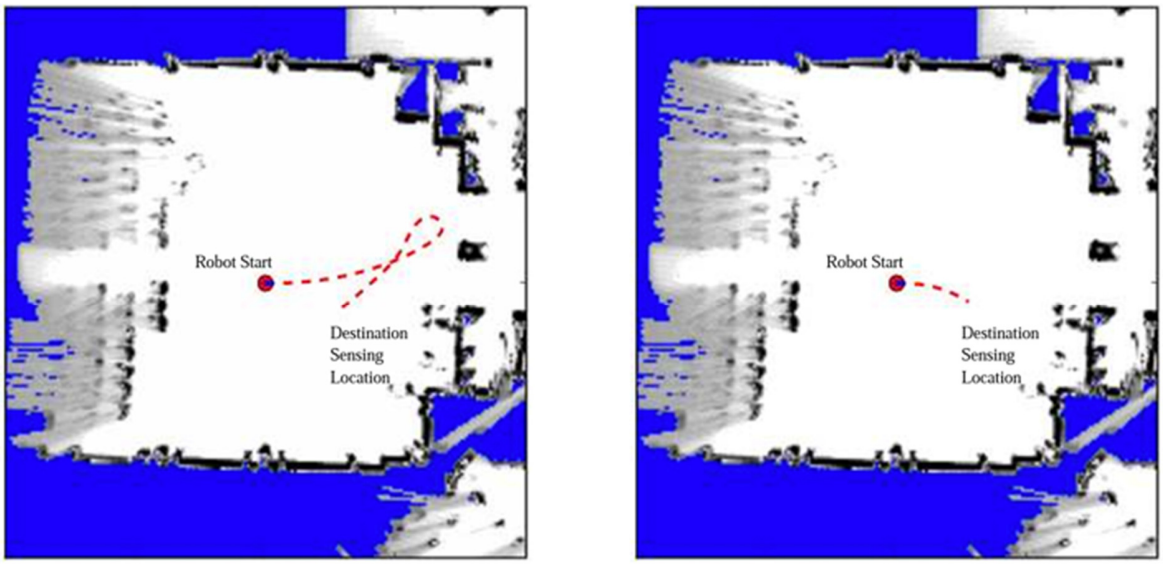
This section will review three topics in the literature using RL methods in path planning and compare them with the proposed model.

A. Trajectory Optimization Using RL

Trajectory optimization is a common problem in the RL field, and successful results have been obtained [8]. Fig. 3 shows two trajectories. However, in these attempts, safety and efficiency are not considered while generating the optimal path, and neither does the user have preferences (like smooth road conditions). So, the exploring algorithm will find the shortest trajectory without considering realistic and essential features such as safety.

B. Q-Learning for Path Planning

Q-learning is a popular algorithm for path planning due to its self-learning without requiring a priori model of the environment, and eventually has been developed to accelerate its performance [9], [10].



(a) Worst Trajectory (b) Best Trajectory

Fig. 3 Example of worst trajectory and best trajectory [8]

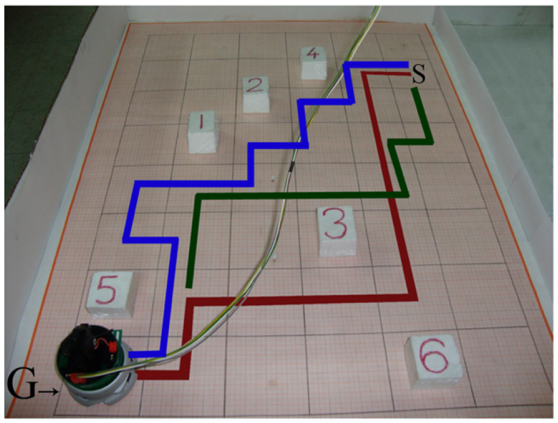


Fig. 4 Results of different types of Q-learning for path planning control [9]

However, these techniques consider two significant states in path planning: "start state" and "goal state." That means the agent (which can be a robot, a drone, or a car) will start its position from the "start state" and proceed to the end destination "goal state." In the middle, there might be "obstacles" that the agent will avoid altogether, and it does not matter what the obstacles are such as a bump or a light standard. However, with the proposed model, the agent differentiates between obstacles, and does not maneuver a specific position unless it is highly unsafe or has a high efficiency cost.

C. Geometric RL for Path Planning

Newly developed RL methods in the literature assist with path planning while enhancing safety. For example in [11], the geometric RL algorithm is used for Unmanned Aerial Vehicles (UAV). It effectively produced good results by allowing a drone to avoid dangerous areas, where dangerous areas could include towering structures or electrical transformer stations.

$$A_{p1,p2} = d_{p1,p2} + K \int_{c_1}^{c_2} F(x,y)\, ds \quad (1)$$

Equation (1) is a part of a developed algorithm called geometric reinforcement learning (GRL), where C is the point set on the path from p1 to p2. dp1 and dp2 are the distance between p1 and p2, and K is the size of threaten parameter that influences the weight between two locations [11]. As K increases the agent will have smaller risk by increasing the distance from the unsafe region. Fig. 5 shows some of the results of the K parameter's effect. In general, their procedure is described as follows:

1. Randomly select start point prm, and m ⊂ Smap
2. Randomly choose target point prn, with n ⊂SPT Pr1
3. Then calculate Aprm,prn as ($A_{prm,prn} = d_{prm,prn} + K \int_{c_1}^{c_2} F(x,y)\, ds$) to get A matrix.
4. Repeat above steps 1–3 until the learning is complete.

The agent (drone) successfully avoids the unsafe zone, but all the unsafe zones have the same level of risk. However, in the proposed model, the risk level can vary from zone to zone, and the agent will respond differently based on the zone's risk level.

III. Background

A. Reinforcement Learning

RL, a general class of Machine Learning (ML), is highly influenced by the theory of Markov Decision Processes (MDP) [12]. Also, RL can be described as a feedback-based ML approach in which an agent is trained on how to behave in a given environment by executing actions and observing the outcomes of those actions, as the agent's goal is to maximize the long-term cumulative rewards.

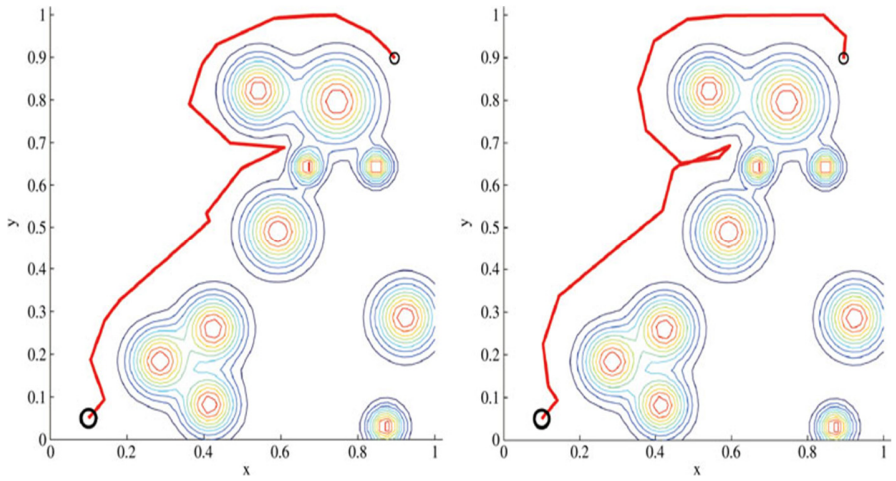


Fig. 5 Different values of the threaten parameter (K) left result with K=20 and the right result with K=100 [11]

General Terms in RL

- Agent: Component that is trained to make the optimal decisions.
- Environment: The system/plant that the agent interacts with as the environment dynamically changes based on the agent's actions and environment's nature.
- Action (a_t): The decision that the RL made to change the environment states.
- State (s_t): The representation or the position of the environment.
- Reward (r_t): A response from the environment that the agent received to assess its performance.
- Policy (π): The strategy that the agent follows when making decisions.
- Q-value ($q(s,a)$): The estimation of the expected cumulative reward of taking a particular action in a given state.

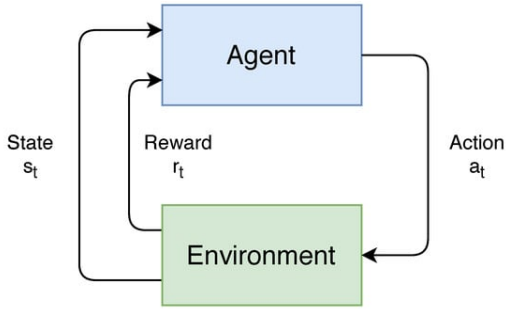


Fig. 6 The RL framework [13]

B. Q-Learning Algorithm

Q-learning methods train the Q-function by iteratively applying Bellman-optimality [14], see (2):

$$q_\pi(s,a) = E[R_{t+1} + \gamma \max_{a'} q_\pi(s',a')] \quad (2)$$

Q-learning is an off-policy algorithm [14], which lets the proposed model achieve the most feasible result while not being bound by a policy that may not allow for the same optimization level. Moreover, it allows for more exploration of a given map. In addition, it is a model-free algorithm [15], so it does not require prior knowledge about an environment, as Q-learning learns about the environment while training and interacting with the environment. Also, this allows the model to be applied to various applications when the underlying dynamics of an environment are unknown. The algorithm is represented in (3) and can refer to the proof of convergence toward the optimality as in [14], [16].

$$Q_t^{new}(s,a) = Q_t(s,a) + \alpha(R(s,a) + \gamma \max_{a'} Q(s',a') - Q_t(s,a)) \quad (3)$$

IV. Model Methodology

The main goal of the model is to guarantee safety and efficiency while choosing the best decision. In this paper, safety can be described as rain or storms (poor weather conditions), a crime scene, or a political protest area, especially if it is unorganized. Efficiency can be represented as EV charge stations along the path, or a less mountainous path. A navigation system was taken to validate and evaluate the proposed model. This approach can thus be used for additional control planning applications such as surgical robots or drone delivery. Therefore, the safety and efficiency criteria of the model can be adjusted based on the application, available data access, and the user's perspective.

A. Model Schematic

The suggested methodology allows the RL model to verify safety and efficiency criteria and, based on the results, adjust the rewards to match the desired performance. As a result, the best policy balances safety and efficiency while making the best decisions. Fig. 7 depicts a high-level overview of the suggested paradigm.

B. Model Functionality

The model will first gather observations from the environment and then store them as numerical data in a reward matrix. Following that, it will check for safety and efficiency standards, and based on the results, it will update the reward matrix. Then, a q-learning matrix will be initialized with zeros

and updated iteratively using the q-learning algorithm until convergence. After creating the q-learning matrix/table, the policy will take the q-learning optimal actions. Fig. 8 depicts a diagram of the model's functionality.

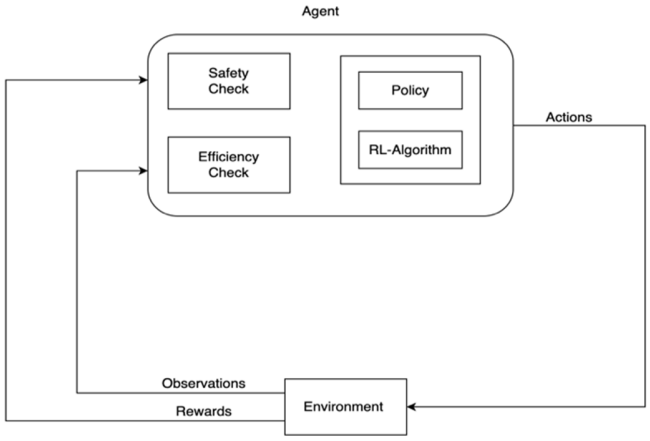


Fig. 7 Model schematic

V. Experimental Setup

The experiment was designed for path planning, going from Tucson, Arizona to Los Angeles, California as an example for the proposed model.

A. Creating the RL - Environment

The first step was examining the best route options using Google Map as in Fig. 9. Then, the change routes have been identified as shown in Fig. 10. Following that, the map was created using Python. Fig. 11 shows the output with the weighted distance from each route change.

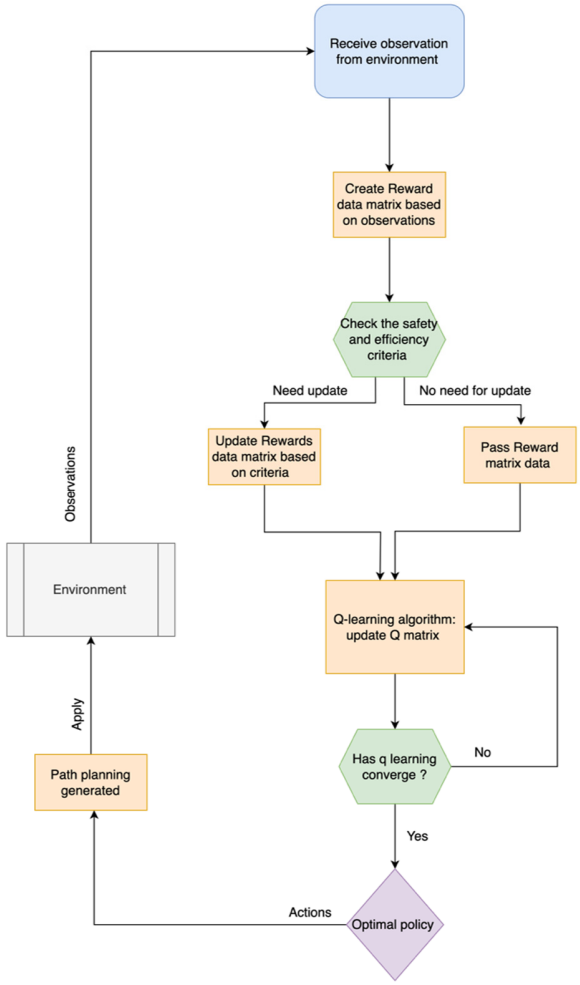


Fig. 8 Model's functionality diagram

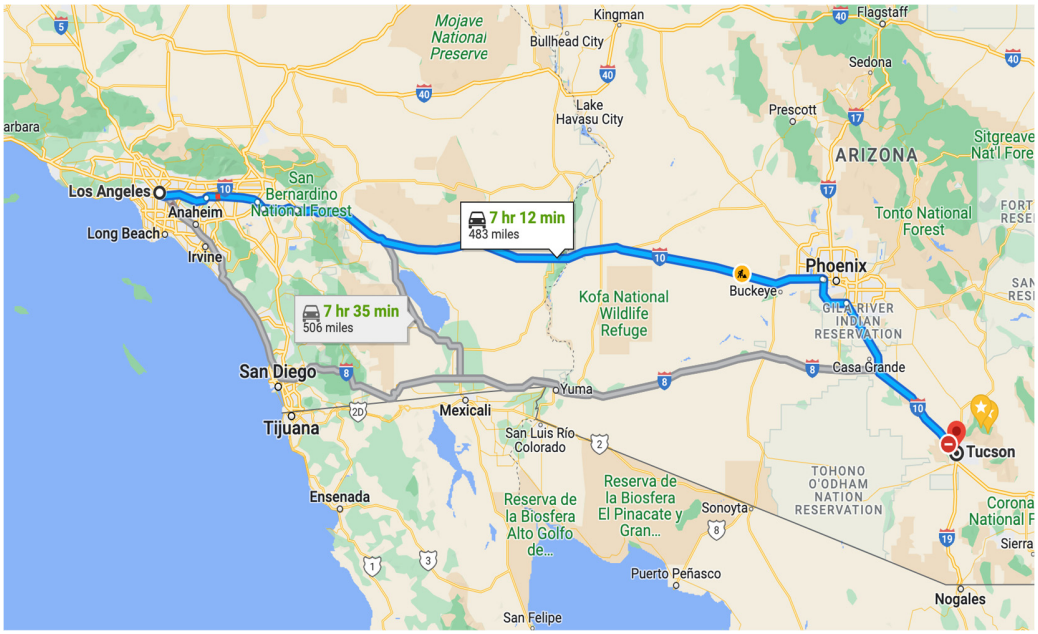


Fig. 9 Best route options from Tucson to Los Angeles

Fig. 10 Identifying change routes

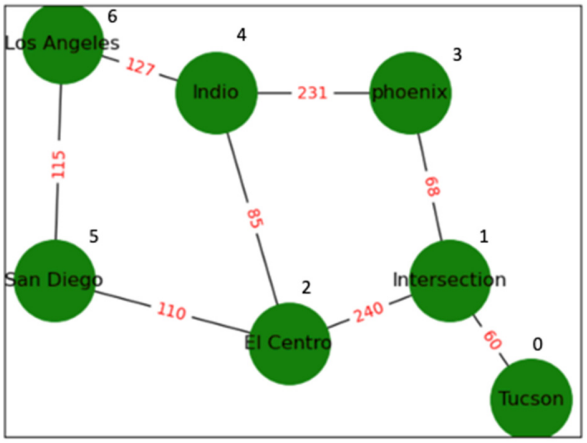


Fig. 11 Output map with distances

B. Training the Agent

The agent was trained via the q-learning algorithm for each run; Fig. 12 is one sample result of 1000 training episodes. The sample shows how the q value increased until convergence.

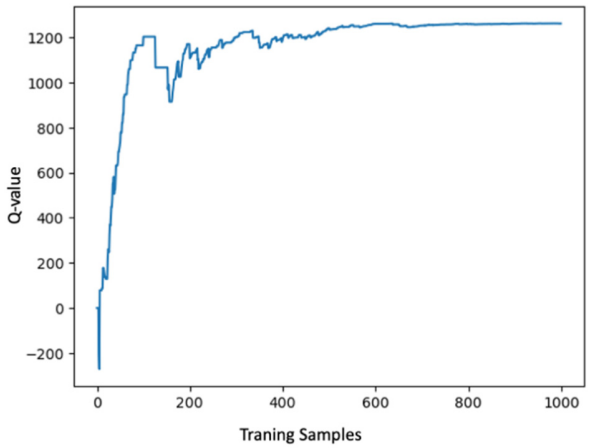


Fig. 12 Training sample

VI. Testing and Evaluating

A. Testing the Model without Safety and Efficiency Features

First, the model was tested without any safety and efficiency checks, and it simply found the shortest path, around 486 miles. the optimal path is colored as blue nodes for clarification in Fig. 13.

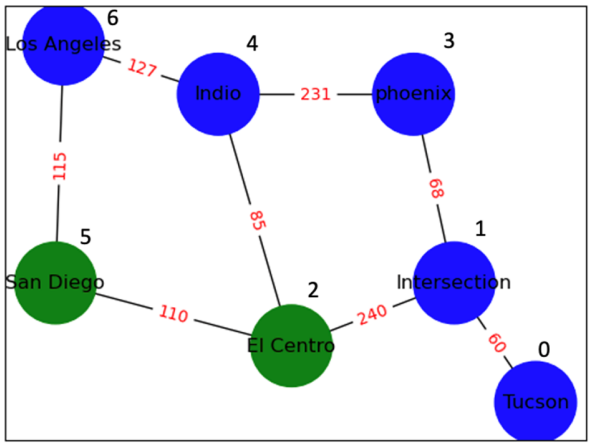


Fig. 13 Performance test without any safety or efficiency checks

TABLE I
Q-Table Of Performance Test without Safety or Efficiency Check

	0	1	2	3	4	5	6
0	0.0000	0.6088	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5419	0.0000	0.6855	0.6896	0.0000	0.0000	0.0000
2	0.0000	0.5906	0.0000	0.0000	0.7887	0.7873	0.0000
3	0.0000	0.6080	0.0000	0.0000	0.7750	0.0000	0.0000
4	0.0000	0.0000	0.6989	0.6731	0.0000	0.0000	0.8871
5	0.0000	0.0000	0.6987	0.0000	0.0000	0.0000	0.8883
6	0.0000	0.0000	0.0000	0.0000	0.7819	0.7878	1.0000

Table I shows how the agent chooses the optimal actions. Table I is a normalized version of the original Q-table for better visualization and analysis, and presents the values as probabilities between 0 and 1. Looking at the table, when the agent is in state 1 (intersection), the agent has three action options: either go back to state 0 (Tucson) by 0.541, go to state 2 (El Centro) by 0.685, or go to state 3 (Phoenix) by 0.689. Therefore, the agent will choose the maximum value that goes to Phoenix. However, the agent did not choose that due to the short distance between the intersection and Phoenix. This means that if the agent only follows the short distance (minimum cost), it would rather go back to Tucson since it has 60 miles rather than 68 miles (going to Phoenix). Indeed, q-

table values are based on the cumulative long-term reward toward the goal state 6 (Los Angeles).

B. Testing the Model with Safety Feature

The model in this section was examined by assuming there is a rainstorm in Phoenix, as the model showed a robust and resilient result by avoiding the rainstorm region (unsafe zone) and following the shortest path. So, when the safety check is enabled, it updates the reward matrix with negative rewards of the undesired place, as the policy follows the desired performance. Fig. 14 shows the result.

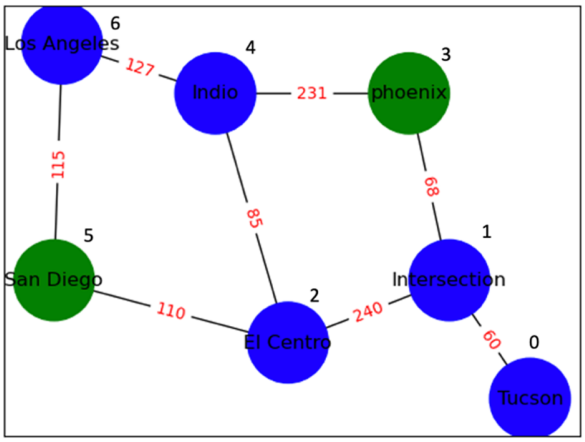


Fig. 14 Performance test with safety check

Table II showed how the agent chose the best action by going to El Centro with a q value of 0.685 rather than Phoenix, whose the q value decreased from 0.689 to 0.670 due to the safety check.

TABLE II
Q-TABLE OF PERFORMANCE TEST WITH SAFETY CHECK ONLY

	0	1	2	3	4	5	6
0	0.0000	0.6107	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5436	0.0000	0.6853	0.6703	0.0000	0.0000	0.0000
2	0.0000	0.5911	0.0000	0.0000	0.7884	0.7851	0.0000
3	0.0000	0.5898	0.0000	0.0000	0.7636	0.0000	0.0000
4	0.0000	0.0000	0.7010	0.6438	0.0000	0.0000	0.8871
5	0.0000	0.0000	0.6984	0.0000	0.0000	0.0000	0.8876
6	0.0000	0.0000	0.0000	0.0000	0.7841	0.7846	1.0000

C. Testing the Model with Safety and Efficiency Features

The model in this section was examined by enabling an efficiency check, assuming the driver is using an electric vehicle and there are EV charge stations in San Diego and none in El Centro and Indio. The model gave good results by following the efficiency desires with the safest path (avoiding the rainstorm in Phoenix and passing through San Diego for the EV charge). Fig. 15 shows the optimal path.

Note that in state 2 (El Centro), the agent chooses to pass through San Diego (taking action 5) by 0.806 rather than going through Indio (taking action 4) by 0.789. Therefore, the San Diego q value increased due to the efficiency check.

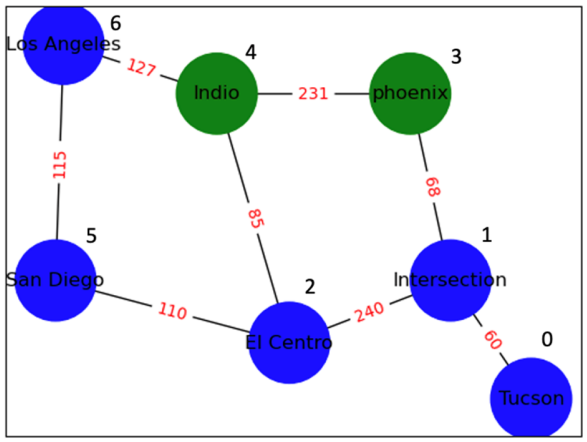


Fig. 15 Performance test with safety and efficiency check

TABLE III
Q-TABLE OF PERFORMANCE TEST WITH SAFETY AND EFFICIENCY CHECK

	0	1	2	3	4	5	6
0	0.0000	0.6256	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5570	0.0000	0.7019	0.6716	0.0000	0.0000	0.0000
2	0.0000	0.6075	0.0000	0.0000	0.7898	0.8068	0.0000
3	0.0000	0.6047	0.0000	0.0000	0.7650	0.0000	0.0000
4	0.0000	0.0000	0.7175	0.6450	0.0000	0.0000	0.8871
5	0.0000	0.0000	0.7352	0.0000	0.0000	0.0000	0.8984
6	0.0000	0.0000	0.0000	0.0000	0.7856	0.8109	1.0000

VII. DIFFERENT LEVEL VALUES OF RISK AND EFFICIENCY

In this section, the model will be demonstrated with different levels of risk and efficiency to show its robustness, and it can be adjusted to be more applicable toward its applied application.

A. Low Risk Level

An assumption is that there is a drizzle in Indio (lower risk) with a risk value of -20 and rain in El Centro (higher risk) with a risk value of -30. Moreover, there are EV charge stations in San Diego with a positive value of 30. Due to the efficiency value, the model let the agent plan to pass through El Centro, taking the rain risk for the efficiency rewards. Fig. 16 and Table IV show the results.

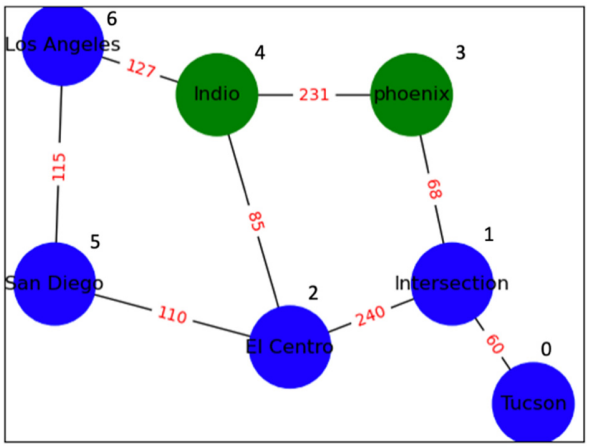


Fig. 16 Performance test with different risk and efficiency level

TABLE IV
Q-TABLE OF PERFORMANCE TEST WITH RISK TAKING FOR EFFICIENCY REWARDS

	0	1	2	3	4	5	6
0	0.0000	0.6101	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5431	0.0000	0.6847	0.6843	0.0000	0.0000	0.0000
2	0.0000	0.5859	0.0000	0.0000	0.7829	0.7911	0.0000
3	0.0000	0.6089	0.0000	0.0000	0.7712	0.0000	0.0000
4	0.0000	0.0000	0.6953	0.6637	0.0000	0.0000	0.8851
5	0.0000	0.0000	0.7039	0.0000	0.0000	0.0000	0.8914
6	0.0000	0.0000	0.0000	0.0000	0.7797	0.7934	1.0000

Looking at the Intersection (state 1), the q-value for choosing Phoenix (action 3) and El Centro (action 2) was very close.

B. High Risk Level

The level of risk in El Centro was increased from -30 to -50, and other parameters were kept as before. As Fig. 17 and Table V show, the model decided to go through Phoenix (q value of 0.687), passing through the low level of risk in Indio while avoiding the high risk in El Centro. In addition, the model neglected the efficiency rewards (EV stations) in San Diego due to the high risk in El Centro as the goal state (final destination) is going to Los Angeles.

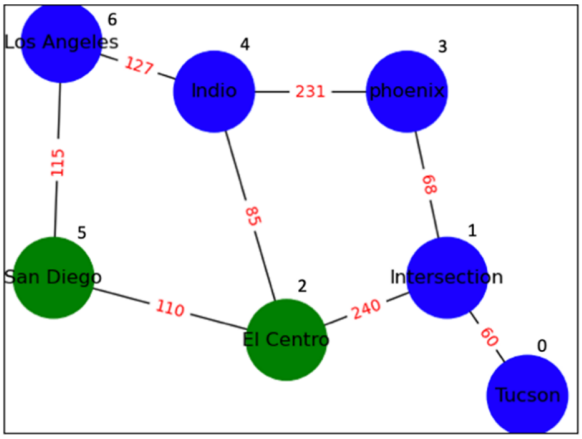


Fig. 17 Performance Test with Different Risk and Efficiency Level

TABLE V
Q-TABLE OF PERFORMANCE TEST WITH HIGH-RISK AVOIDING AND NEGLECTING EFFICIENCY REWARDS

	0	1	2	3	4	5	6
0	0.0000	0.6124	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5437	0.0000	0.6808	0.6871	0.0000	0.0000	0.0000
2	0.0000	0.5827	0.0000	0.0000	0.7787	0.7888	0.0000
3	0.0000	0.6100	0.0000	0.0000	0.7715	0.0000	0.0000
4	0.0000	0.0000	0.6876	0.6668	0.0000	0.0000	0.8852
5	0.0000	0.0000	0.6999	0.0000	0.0000	0.0000	0.8914
6	0.0000	0.0000	0.0000	0.0000	0.7795	0.7955	1.0000

VIII. CONCLUSION AND FUTURE RESEARCH

This paper presents an RL control model for path planning applications that can be used to improve safety and efficiency. The navigation system was provided to illustrate the model's concept and techniques.

Other methods in the literature, such as in [8], improved path planning, but safety and efficiency were not considered. Nevertheless, the presented model showed the robustness of making decisions following the optimal path, while considering safety and efficiency. Also, in the literature, for instance, in [9], [10], or [11], these approaches improved safety for path planning but assumed the zone risks had the same levels for all the unsafe regions. However, with this model, the safety and efficiency values can vary depending on the risk and efficiency situations. The presented model can compare risk levels and efficiency advantages and consider that intelligently while making its decision.

This RL control model can incorporate many feature checks. For example, if there is a need to access hospitals along the way, avoid mountainous terrain, or other driver dependent features. The poor weather conditions and EV charge stations were taken to show the model response for safety and efficiency. The results show that the agent successfully avoided the unsafe regions, while considering efficiency advantages.

As a limitation of this research, the model is not a replacement for the navigation system. Indeed, it is an added bonus to improve navigation systems and other path-planning control applications to acknowledge other features while generating optimal paths.

In future research, the model can be applied to other control planning applications like surgical robotics, varying the risk level between human organs, as the agent should take feature checks under consideration while making its decisions. In addition, the model could be applied to physical systems to show how the agent intelligently takes responsibility for other matters, not just minimizing cost or arriving faster.

REFERENCES

[1] Nambiar, Kavya. "How Do Google Maps Work?" Analytics Steps, 6 June 2021, www.analyticssteps.com/blogs/how-do-google-maps-work. Accessed 18 Sep. 2023.
[2] Castrodale, Jelisa. "This Is How Google Maps Knows Which Route Is the Fastest at Any given Moment." USA Today, 24 Nov. 2015, www.usatoday.com/story/travel/roadwarriorvoices/2015/11/24/this-is-how-google-maps-knows-which-route-is-the-fastest-at-any-given-moment/83282460/. Accessed 27 Sep. 2023.
[3] Federal Highway Administration (FHWA), U.S. Department of Transportation (DOT). "How Do Weather Events Impact Roads?", 1 Feb. 2023, https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. Accessed 2 Oct. 2023.
[4] Pelegov, Dmitry V., and Jean-Jacques Chanaron. "Electric Car Market Analysis Using Open Data: Sales, Volatility Assessment, and Forecasting." Sustainability, vol. 15, no. 1, 2022, p. 399, doi:10.3390/su15010399.
[5] International Energy Agency (IEA). "Electric Car Sales, 2016-2023 – Charts – Data & Statistics." 26 Apr. 2023, www.iea.org/data-and-statistics/charts/electric-car-sales-2016-2023. Accessed 5 Oct. 2023.
[6] DriveClean.Ca.gov "Electric Car Charging Overview", 2021. driveclean.ca.gov/electric-car-charging. Accessed 11 Oct. 2023.
[7] Pacey, Paddy. "Honey Buzzard – Epic Migration – WildAware Environmental Conservation Organization." Whole Earth Education, 20 Oct. 2020, wholeeartheducation.com/honey-buzzard-epic-migration/. Accessed 13 Oct. 2023.
[8] Kollar T, Roy N. "Trajectory Optimization using Reinforcement Learning for Map Exploration.", The International Journal of Robotics Research. (2008) 27 (2):175-196.
[9] Konar, Amit, et al. "A deterministic improved Q-learning for path planning of a mobile robot." IEEE Transactions on Systems, Man, and Cybernetics: Systems 43.5 (2013): 1141-1153.
[10] Low, Ee Soong, Pauline Ong, and Kah Chun Cheah. "Solving the optimal

path planning of a mobile robot using improved Q-learning." Robotics and Autonomous Systems 115 (2019): 143-161.
[11] Zhang, B., Mao, Z., Liu, W. et al. “Geometric Reinforcement Learning for Path Planning of UAVs.” J Intell Robot Syst 77, 391–409 (2015).
[12] Hutsebaut-Buysse, M.; Mets, K.; Latré, S. “Hierarchical Reinforcement Learning: A Survey and Open Research Challenges.”, Machine Learning and Knowledge Extraction, vol. 4, no. 1, 2022, pp. 172–221.
[13] Wiering, Marco A., and Martijn Van Otterlo. "Reinforcement learning." Adaptation, learning, and optimization 12.3 (2012): 729.
[14] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8 (1992): 279-292.
[15] Glorennec, Pierre Yves. "Reinforcement learning: An overview." Proceedings European Symposium on Intelligent Techniques (ESIT-00), Aachen, Germany. 2000.
[16] Watkins, Christopher John Cornish Hellaby. "Learning from delayed rewards." (1989)