Certification of Reinforcement Learning Applications for Air Transport Operations Based on Criticality and Autonomy

Ribeiro, M.J.; Tseremoglou, I.; Santos, Bruno F.

DOI
10.2514/6.2024-1463

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the AIAA SCITECH 2024 Forum

AIAA SciTech Forum
8-12 January 2024, Orlando, FL
AIAA SCITECH 2024 Forum

10.2514/6.2024-1463

Certification of Reinforcement Learning Applications for Air Transport Operations Based on Criticality and Autonomy

Marta Ribeiro *, Iordanis Tseremoglou†, and Bruno F. Santos ‡
Delft University of Technology, 2629 HS Delft, The Netherlands

Despite its success in various research domains, Reinforcement Learning (RL) faces challenges in its application to air transport operations due to the rigorous certification standards of the aviation industry. The existing regulatory framework fails to provide adequate, acceptable means of compliance for RL applications, and thus, there is no legal framework for their safe deployment yet. Guidelines must be formulated to certify RL models aimed at air transport operations to enable real-world utilisation of these promising methods. These guidelines must consider the unique characteristics of these models, deviating from the methodology of current guidelines crafted before the emergence of ML applications. The paper proposes novel certification requirements for RL models based on their technical characteristics, safety-criticality, and autonomy. This framework covers the choice of the RL algorithm and analyses the actions, agents, environment, and potential hazards and risks of the RL application. Additionally, this work outlines the evidence the certification applicant must present to demonstrate compliance with these requirements. While this framework is not a complete solution for the complex problem of certifying RL, it is intended to serve as an initial framework which can be extended in cooperation with regulatory entities.

I. Introduction

The field of reinforcement learning (RL), a subset of machine learning (ML), has received much interest from many research fields and is now widely used in operations research. Applications vary greatly, from robotics and computer systems to finance and healthcare [1]. Nevertheless, air transport operations have not followed this trend. The fact that airborne applications must be developed in compliance with the rigorous certification standards of the aviation industry explains the slow progress in the use of RL approaches in this domain.

However, there is a growing interest in the potential of RL to handle emergent behaviour in air transportation [2, 3]. This emergence results from the interacting dynamics between multiple aircraft, airlines, and airports [4]. The final result of this large system cannot be explained by looking at the individual properties of each agent. Emergence can have adverse effects when not properly understood. Nevertheless, RL models can adapt to this emergent behaviour by training directly in the operational environment. Recent promising RL solutions in aviation include autonomous separation assurance in Air Traffic Management (ATM) [5–8], and decision-making on predictive maintenance [9, 10]. However, one of the biggest hurdles to these applications is how to certify them - for which there is no answer yet.

Certification standards for the aviation industry were developed before the emergence of ML applications without considering the specifics of these models. As a result, the current certification is not directly applicable to ML. Furthermore, the former focuses heavily on the predictability and explainability of the models; these are understandably crucial factors for end-user acceptance. However, many works fail to explain the RL model's results and decisions. This is not the researchers' fault but rather a shortcoming of these models. Some examples of initial research attempts to circumvent the lack of explainability in ML [11, 12]. Nevertheless, these often focus only on supervised learning. The latter predicts future behaviour learning directly from available data; thus, behaviour can be extrapolated from the input data. RL is an entirely different mechanism - the optimal actions the RL agent finds are often unpredictable, found through interaction with the environment alone.

Moreover, the most promising RL applications employ deep neural networks (i.e., deep reinforcement learning (DRL)) [13]. These applications are more demanding to certify, given their limited explainability —- actions from DRL result from a combination of a high number of weights. The combination is the defining factor, not the weights themselves - a difficult concept for humans to grasp. Consequently, it is still unclear how this research can reach

*Assistant Professor, Control & Simulation Department, Faculty of Aerospace Engineering, M.J.Ribeiro@tudelft.nl
†PhD Candidate, Control & Simulation Department, Faculty of Aerospace Engineering, I.Tseremoglou@tudelft.nl
‡Associate Professor, Control & Simulation Department, Faculty of Aerospace Engineering, B.F.Santos@tudelft.nl

end users, although there is a lot of interest in making this happen, given their potential. Recent work highlights the incompatibilities between the conventional design assurance approach and aspects of ML systems [14, 15]. In early 2023, the European Union Aviation Safety Agency (EASA) published the concept paper: 'First usable guidance for Level 1&2 machine learning applications' [16], which aims at providing guidelines for applicants introducing ML to safety-related or environment-related applications in the aviation domain. Nevertheless, this document is still in the initial phase and is directed only at supervised learning (extensions of this document toward RL and unsupervised learning are planned for the future). This work is directed at this certification gap.

This paper innovates by proposing novel certification guidelines for RL models aimed at air transport operations. The lack of such guidelines is the main factor hindering the development of these models and their application by end-users. In summary, the objectives of this paper are as follows:

1) To highlight the benefit of introducing RL applications in air transport operations.
2) To give an overview of the existing regulatory framework for certifying RL-based systems.
3) To propose novel certification requirements for RL models based on their technical characteristics, safety-criticality, and autonomy.
4) To define the evidence the certification applicant must present to prove compliance.

The paper is structured as follows: Section II presents an overview of current RL applications in aviation. Emphasis is placed on the differences between these operations regarding safety criticality and autonomy (e.g., an application directed at aircraft control versus decision support solutions). Section III introduces the regulatory framework for applying RL in air transport operations. The current guidelines from EASA and FAA are covered. In addition, this section covers research efforts focused on the certification of ML in aviation beyond the existing directives of regulatory bodies. The proposed RL certification methodology is presented in Section IV. This section starts with an analysis of the technical foundations of RL, highlighting the differences between these models and the traditional software applications covered by current certification rules. The unique characteristics of RL form the cornerstone of the proposed certification framework. Section V discusses future steps to improve this novel framework. Finally, Section VI brings this paper to a close.

II. Applications of Reinforcement Learning in Aviation

Table 1 presents an overview of the diverse areas within aviation where RL is actively applied. These areas are evaluated regarding criticality and autonomy in alignment with the framework proposed in this work (Section IV). The level of safety-criticality pertains to the safety requirements of the application domain of the RL algorithm. An RL framework that controls the aircraft's path and attitude or mitigates the potential loss of separation events is attributed to the highest criticality (++). On the other hand, autonomy indicates the degree to which the RL application acts as an auxiliary decision support system (non-autonomous (+)) or assumes full decision control, circumventing human intervention (autonomous system (++)). Note that, in some cases, both decision support and autonomous applications can be developed for the same area. For instance, in autonomous taxiing, RL may be used to predict taxiing time [17] or to direct the path of the aircraft [18].

Previous research efforts have extensively investigated the application of RL in aviation [2, 3]. These studies highlight the growing volume of research and the critical need for dedicated attention to certify such innovative applications in aviation. Additionally, researchers argue that the current lack of certification hinders advancements in the field, as there may be considerable differences between the simulated environment and real-world applications. However, a more concrete analysis of these differences and how to improve the realism of these methods is hindered by the absence of a legal framework permitting their deployment in real-life operations.

Finally, it's worth noting that the studies encompassed in Table 1 exclude research for future urban air mobility (UAM) applications. This paper focuses on current operations with a technology readiness level that supports their certification. Nevertheless, the rising number of RL studies in the realm of UAM is notable, as indicated by recent studies [31]. This interest likely results from the need to automate these pilotless, on-demand services. For reference, EASA's 'Artificial Intelligence Roadmap 2.0' [15] mentions that *'U-space services will only be possible with high levels of automation and use of disruptive technologies like artificial intelligence (AI) and ML'*, especially in the areas of detection and avoidance solutions and autonomous navigation.

Table 1 Works employing reinforcement learning in the field of aviation.

Area of Research	Level of Safety-Criticality	Level of Autonomy	Exemplary Works
Aircraft Maintenance/Failure Prediction	+/++	+/++	[9, 19, 20]
Autonomous Air Traffic Control	++	+	[5, 6, 21]
Autonomous Taxiing	++	+/++	[17, 18]
Attitude Control	++	+	[8, 22]
Crew Allocation	+	+/++	[23]
Flight Delay Prediction	+	+	[17]
Path Planning	++	+	[24, 25]
Pilot Behaviour Prediction			[26, 27]
Revenue Management		+	[28]
Runway Scheduling	+	+	[29, 30]

III. Current Regulatory Framework

As mentioned previously, the regulatory landscape for certifying RL in aviation is severely limited. This section presents an overview of the existing directives, including potentially applicable ML regulations. The existing regulation is as follows:

- EASA's 'First usable guidance for Level 1 & 2 machine learning applications' [16]: document intended as an early visibility on the possible expectations of EASA for future approval of AI/ML applications. Nevertheless, this document is focused on supervised learning and does not cover RL.
- EASA 'Concept of Design Assurance for Neural Networks' [32]: this document aims to provide users with understandable and relevant information on how AI/ML applications decide upon actions. This assumes that the distribution of inputs in the learning data is correct and, thus, is only applicable to controlled, supervised learning models.
- FAA's Certification Research Plan for Artificial Intelligence Applications [33]: plan to create an initial certification framework. This plan mentions: *'While the potential benefits of AI and ML are clear, the evolving, non-deterministic nature of these algorithms presents challenges for the air traffic management industry, aircraft, and new entrants (e.g., Unmanned Aircraft System) regarding the certification of these algorithms'*.
- NASA's 'Challenges in the Verification of RL Algorithms' [34]: document aimed as a first step towards verification and adoption of RL in autonomous systems. This paper proposes verifying state-to-action mappings and action sequences, monitoring incorrect actions, and detecting situations not present in the training data.

Given the scarcity of frameworks from regulatory entities, researchers are actively working to bridge this gap to facilitate the integration of RL into real-life operations. Noteworthy works resulting from this effort are herein mentioned:

- Baheri et al. [35]: proposes a module for monitoring potential unsafe actions from an RL application directed at path control. The module alerts the human in case an unsafe action is detected.
- Dmitriev et al. [14]: work aimed at showing that current airborne certification standards can be complied with, in the case of a low-criticality ML-based system, if certain assumptions about the development workflow are applied. Nevertheless, these assumptions are only applicable to supervised learning systems.
- Torens et al. [36]: detailed overview of the existing regulatory framework for certifying AI-based systems in aviation, noting that it fails to provide adequate acceptable means of compliance. In response to this gap, the paper proposes to look into standardisation frameworks from other domains, specifically the automotive one.
- DEEL's project [37]: ongoing project to create guidelines for the explainability and robustness of AI algorithms. This project highlights how understanding the decisions of ML models plays a key role in achieving certification. However, it is restricted to the analysis of offline supervised learning techniques.
- DARPA's Explainable Artificial Intelligence Programme [38] (2015-2021): with the goal of enabling end users to better understand, trust, and effectively manage AI systems. Lessons learned from the project highlight that users prefer systems that provide decisions with explanations over systems that provide only decisions.

In summary, the certification of RL is still in its early stages, both for the regulatory entities and researchers. The possibility for certification of ML in aviation currently covers only specific implementations of supervised learning, when it is trained with verified data. The more RL-oriented works [34, 35] recommend verification of whether actions output by the RL model align with predefined ranges of safe or permissible actions. Note that the present paper includes this aspect and proposes additional elements for verification.

IV. Proposed Methodology

We identify four general RL classification levels to provide a safety-oriented framework for our discussion. This scheme is proposed based on the (1) safety criticality of the RL application and (2) the authority of the RL model vs the end user, ranging from full or partial authority of the end user (decision-support system) up to full authority for the RL-based model (autonomous system). Taking into account the former two criteria, and to promote consistency and better understanding between the aviation stakeholders, we name and prioritise (from A-higher priority to D-lower priority) the four RL classification levels considering the consequences related to potential hazards (taking into account the worst foreseeable situation) following the typical safety risk severity table as defined in ICAO Doc 9859 [39], i.e.:

Level A: Catastrophic - prevents continued safe operations or control and could lead to many fatal injuries.
Level B: Hazardous or severe - prevents continued safe operations or control and could potentially lead to fatal injuries to people.
Level C: Major - impairs operations efficiency, discomfort, or injuries to a few people.
Level D: Minor - reduced safety margins, which the human operator or controller can cover.

Figure 1 depicts the resulting classification scheme and provides a reference for classifying the RL-based models. A higher classification level requires a more extensive and meticulous certification compliance process. Also, the expected level of robustness of the RL model increases with the operation risk. The following sections define different certification compliance rules according to the RL classification level. Tables 3 to 7 display the requirements for the certification applicant. Note that, for each level, the applicant must comply with the requirements up to that level. For example, for Level C, the applicant must comply with both Levels D and C; for Level A, the applicant must comply with the requirements for Levels D, C, B and A. Moreover, in case of doubt, the applicant should assume the higher priority classification level.

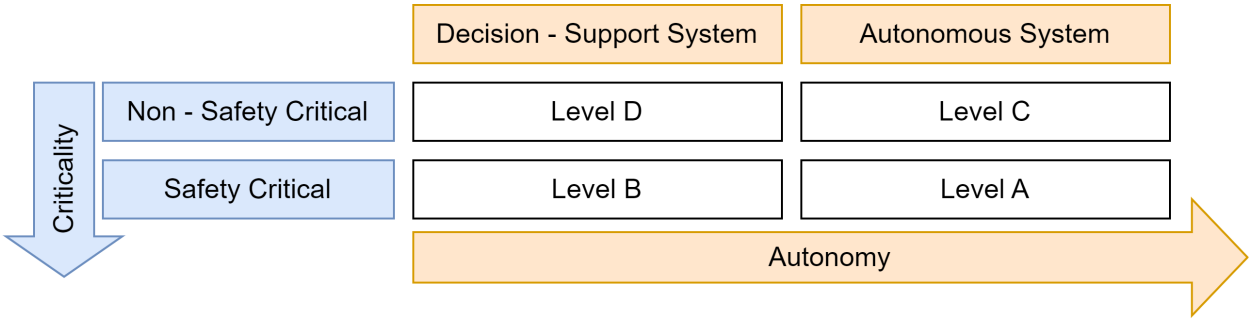

Fig. 1 The different RL classification levels are considered in this paper. These are divided according to the autonomy and safety criticality of the final system in which the RL model is deployed.

A. Overview of the Certification Elements Composing this Framework

This certification framework focuses on the main elements of training, testing, and verification of an RL model, as shown in Figure 2. The RL model is composed of an RL algorithm to which it is assigned a set of inputs (i.e., defining the current state of the environment) and a set of outputs (i.e., the actions to be performed by the agent) that result in a change of the state of the environment. This change is evaluated through the means of a reward assigned to the model. This reward is used as reinforcement and updates the algorithm's expected reward for the action taken. The algorithm aims to find which actions produce environmental changes that maximise the reward. An RL algorithm learns by trial and error by registering the reward received for every state-action pair.

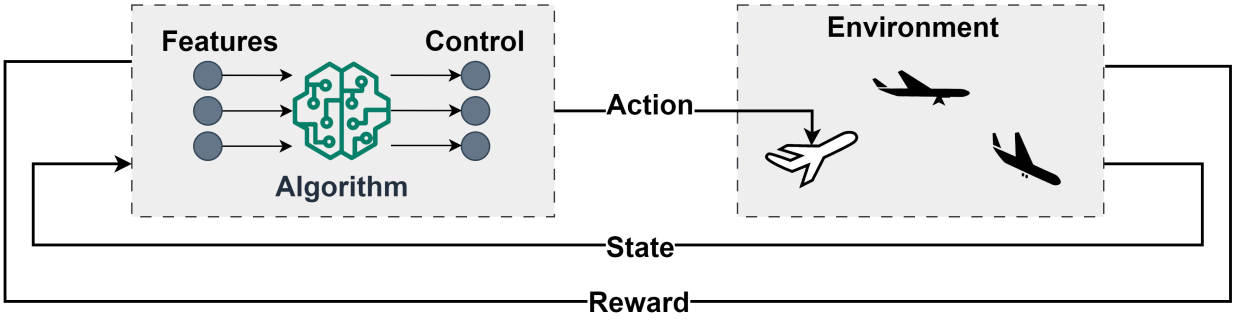

Fig. 2 Overview of the several components involved in the training, testing, and verification of an RL model.

The following areas are considered to have the most impact on the final safety assessment of the RL model. They are, therefore, the areas involved in the certification framework proposed in this paper:

- The RL algorithm (Section IV.B): covers the type of algorithm employed. An inverse relationship exists between scalability to larger state/action spaces and explainability/predictability. RL algorithms that can scale to larger state/action spaces have less explainability/predictability. The choice of which RL algorithm to pick must be reasoned based on the final desired functionality, the space/action space size, and the user requirements on explainability and predictability.
- The (simulated) environment (Section IV.C): the RL model learns which actions result in desired changes to the environment. These changes must correctly parallel the reactions of the final deployment scenario to guarantee that only safe, desirable actions are performed by the model.
- The (simulated) agent (Section IV.D): the RL model is responsible for defining the actions to be performed by a specific agent. A degree of similarity between the agent during training and the final operating agent must be established. Otherwise, it cannot be guaranteed that the actions of the RL model will be optimal.
- The (simulated) actions (Section IV.E): the RL agent should not output non-allowed actions in the deployment environment. Additionally, the sequence of actions must not place the environment in an undesirable state.
- Hazard and risk modelling (Section IV.F): the potential risk and worst-case scenarios resulting from the deployment of the RL model must be clearly defined. Safeguard techniques are discussed to prevent the RL model from moving the agent/environment towards undesirable, non-safe states.
- Update of the RL model (Section IV.G): RL models may be continuously improved through new data. Nevertheless, re-training an RL model, contrary to other methods, may not result in a simple model extension. It can, instead, redefine the weights that the model uses to relate the received input and the produced actions. As a result, we are faced with a new model which must be again certified if employed in safety-critical areas.

B. RL Algorithms - Explainability & Predictability

This section covers the rules for certifying RL applications regarding their explainability and predictability. The former increases the user's trust in the final RL application and is a strong measure for verification that the application performs as expected. The latter guarantees that the application will continue to perform within the limits covered in its certification. This framework distinguishes between two different categories of model-free RL algorithms, in line with literature [40]:

- Tabular Solution Methods (TSM): in which the state and action spaces are small enough for the approximate value functions to be represented as arrays or tables (e.g., discrete action/state spaces). In this case, the methods can often find exact solutions - they can often find exactly the optimal value function and the optimal policy. Examples of these methods are as follows:
  - Multi-armed Bandits
  - Monte Carlo Methods
  - Finite temporal-difference methods (e.g., Q-learning [41], SARSA [42])
- Approximate Solution Methods (ASM): used in problems with arbitrarily large state/action spaces. In this case, the model cannot expect to find an optimal policy or the optimal value function even in the limit of infinite time and data (e.g., continuous action/state spaces). The goal, instead, is to find a good approximate solution using limited computational resources. Examples of these methods are as follows:

– Deep Reinforcement Learning (DRL), where neural networks approximate the environment's dynamics, such as Deep Q-Learning (DQN) [43], and Deep State-Action-Reward-State Action (D-SARSA) [42] algorithms.
– Actor-Critic, where the tasks of estimating both the policy and the value function is done by two agents (actor and critic, respectively), such as Deterministic Policy Gradients (DPG) [44], Advantage Actor-Critic (A2C) [45], and Soft-Actor Critic (SAC) [46] algorithms.

RL models are told *what* to do and find out *how* to do it. However, in safety-critical applications *what* and *how* are equally important. The ability to answer the *what/how* questions refers to the explainability and predictability of the model, respectively. Nevertheless, a negative correlation exists between explainability/predictability and the model's scalability in environments with a large state/action space. An overview of commonly used RL algorithms in relationship to their scalability, explainability, and predictability is shown in Figure 3.

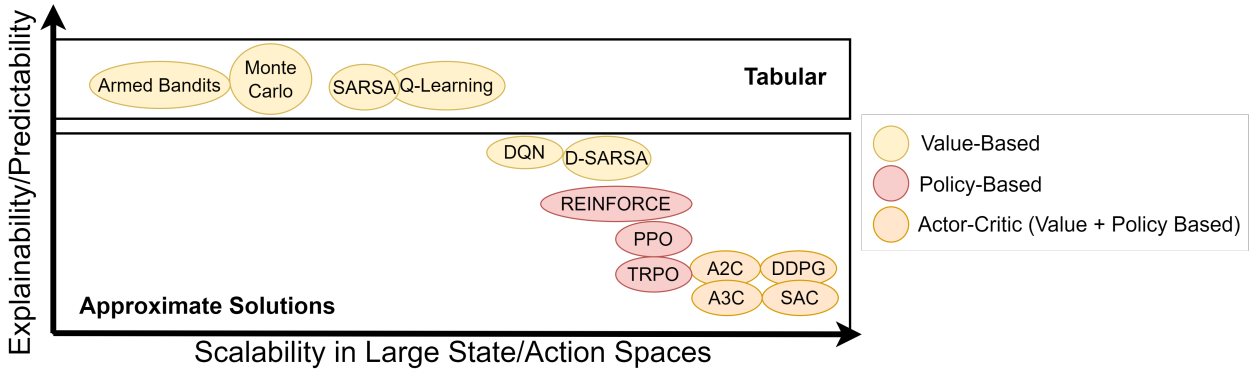

Fig. 3 Scalability of RL algorithms in large action/space spaces versus their explainability/predictability.

The certification applicant must clearly justify the choice for a specific RL algorithm. This justification shall align with the pre-defined requirements for the RL model. The expected state/action space range must be set in advance. In practice, this represents the range of scenarios and actions the RL model is expected to experience and be responsible for in the deployment environment. Methods better suited for large spaces allow a limited understanding of their inner workings. This drawback should only be endorsed when necessary, given the operational scenario and the objectives for the RL method. Methods to potentially explain and predict the actions from different RL algorithms are explored in Table 2 - the extent of the explainability and predictability that can be achieved is limited by the type of RL algorithm.

Table 3 shows the proposed certification compliance requirements for different RL algorithms based on their explainability and predictability limits. The proposed requirements are divided per the previously defined classification levels (i.e., Level D to A).

C. Environment Definition

This section covers the rules for certifying RL applications regarding their training environment. This is considered one of the main elements in this certification framework. Due to the complexity of simulating all the dynamics in a real-world environment, considerable differences exist between the simulated environment and real-world applications. However, assurance of the efficacy/efficiency of the RL model is directly connected to the environment in which the RL application is trained. Once trained in a given environment, it cannot be assumed that the model's behaviour can be directly extended to another environment. The certification applicant must thus clearly declare to which extent the environment in which the RL model was trained mimics the final operational environment in which the RL model is to be deployed.

Note that two different forms of training may be considered: offline and online. In the former, the agent only uses data collected (e.g., from human demonstrations) without interacting directly with the environment. In the latter, the RL agent gathers data directly by interacting with the environment. RL commonly learns online, and thus, the certification rules in this section focus on this type of training. Nevertheless, the certification requirements are still compliant with offline training. Here, the certification applicant must still prove that the data used to train the RL model is a fair and (semi-) complete representation of the RL world.

Table 2 Explainability and predictability limits for both tabular and approximate solutions RL algorithms.

	Tabular Solution Methods	Approximate Solution Methods
Explainability	Can be fully achieved: all states and actions can be fully represented; a direct correlation can be inferred	Cannot be fully achieved: an (almost) infinite number of possible state representations and actions. The relationship between state and action is approximated. Methods to increase explainability: • Feature Importance analysis (i.e., additive feature attribution methods, Shapley method) • Diagrams of Output Actions Per Variation of Main Features
Preditability	Can be fully achieved: it is possible to predict which state leads to each action. All actions are known	Cannot be fully achieved: an (almost) infinite number of possible state representations and actions. The relationship between state and action is approximated. Methods to increase explainability: • Create a diagram of the spectrum of actions given large variations of state values • Run a large testing phase with great variability of cases

Table 3 Certification compliance requirements for the RL algorithm, per classification level.

	Tabular Solution Methods	Approximate Solution Methods
Level D	Demonstration of uses cases proving the success of the decisions of the RL model for the expected operational cases	
Level C	Demonstration of uses cases fairly representing the range of operational cases that the RL model is expected to act upon	
Level B **Level A**	Full analysis of all states and actions and direct correlation between which states lead to specific actions	Full explainability/predictability is not possible: the RL model's application must be limited to previously tested state/action combinations

The applicant should describe the following to show how the simulated environment compares to the operational environment where the RL application is to be deployed:

- All the potential variables, uncertainties, perturbations, or environmental and structural changes at play in the final deployment scenario. Note that, in environments with a high level of unpredicted uncertainty, any trained optimal policy of the RL application could perform poorly when the uncertainty level extends far from the cases modelled during the training of the model.
- The feasible range of different states of the deployment operational scenario, compared to the range of states in the environment in which the model was tested. If the number of possible states of the environment cannot be defined, then the safe range of states to which the RL model can respond must be defined instead.
- The expected changes in the environment result from the actions of the RL application. The latter estimates the transitions of the environment in response to each action. The level of certainty regarding these transitions must be considered. If the environment is to alter in a way not previously experienced by the RL application, its policies will not be optimal.

Table 4 defines the proposed compliance rules for certification, focusing on comparing the environment in which the RL application is trained and the final deployment environment.

D. Agent Definition

This section covers the elements that the certification applicant must define for the efficiency and safety of the actions performed by the RL application to be evaluated. The application learns the optimal actions for a specific agent.

Table 4 Proposed requirements regarding the environment definition, per certification level.

Requirements - Environment Definition	
Level D	The applicant declares that the RL application was trained in a fair representation of the deployment operational environment
Level C	The RL application must be trained in a simulated environment proven to have a good degree of similarity with the deployment operational environment
Level B	Full analysis of all state/actions and direct correlation between which states lead to specific actions. The RL application must be trained in a simulated environment proven to be similar to the deployment operational environment. The main sources of uncertainties and environmental perturbations must be identified, categorised, represented, tested, and validated
Level A	The RL application must be trained in a real operational (safe) environment or a simulated environment with a high level of fidelity to the deployment environment. All possible ranges of uncertainties and environmental perturbations must be identified, categorised, represented, tested, and validated

The responses received by the application are directly related to which actions the RL agent can successfully implement and how fast it can perform them (i.e., the RL application indirectly learns the performance limits of the agent). The model assumes constant performance and response times for the same actions. Consequently, declines in the agent's performance or delayed response in the deployment operational scenario will hinder the model's efficacy. The following must be taken into account:

- The RL application must be trained with a fair representation of the final agent whose actions it will decide once deployed in the operational environment. All assumptions made on the dynamics and performance of the agent should be declared.
- The performance limits (e.g., maximum speeds, turn rate, prediction accuracy) of the agent that the RL application was trained with must fairly represent the performance limits of the final operational agent. It must be examined whether the RL application will output a change in the agent's current state that the latter cannot achieve (e.g., reach a certain speed or position in a certain amount of time).
- Safeguards must be in place to prevent the RL application from acting when the agent does not respond or perform as expected. RL applications are not trained to contemplate changes in the agent's performance. Any communication delays between the RL application and the final operational agent must be examined.
- Multi-agent emergence: multi-agent interactions may lead to unpredictable behaviour. Air transportation is full of resultant emergence [4] - this emergent behaviour is dependent on the number of actions that all agents in the environment can perform, their interactions, and their performance. The expected interactions between the agent controlled by the RL application and other agents in the environment must be clearly defined. Safety in a multi-agent environment is of high complexity and importance - it must consider the safety and efficiency of the controlled agent and the neighbouring agents.

Table 5 displays the certification compliance rules advised regarding the agent that the RL application controls.

E. Action Definition

This section covers the elements that the certification applicant must define for the safety of the actions performed by the RL agent to be evaluated. It must be confirmed, prior to deployment, that the RL application does not perform illegal actions that would put the environment or other agents at risk.

The applicant must clearly define the extent to which the actions of the RL application can position the agent or the environment in a (nearly) fatal situation. Often RL applications may choose a temporary 'bad' state if the model finds that this state can lead to a 'good' state. If there is a possibility that the model can make such decisions, these states must be pre-declared and their risk assessed. For safety-critical applications, (near-)fatal states shall not be allowed - the transition of the RL application must be restricted to 'safe' states. This is denominated as 'safe' RL [47, 48]. However, it must be taken into account that prohibiting certain ('bad') states is penalising exploration, which may reduce the optimality of the actions found in the RL application. The latter samples and explores the state space to improve policy estimation and achieve convergence. It is a tricky balance to allow sufficient exploration while blocking specific states. Finally, determining these unsafe states is a challenging problem, especially when the dynamics of the environment are

Table 5 Proposed requirements regarding the agent definition, per certification level.

Requirements - Agent Definition	
Level D	The applicant declares that the agent the RL application was trained with is a fair representation of the final operational agent
Level C	Use cases are in place that show that the training agent is a fair representation of the final operational agent
Level B	The RL application must be trained with a highly faithful representation of the final operational agent. Safeguards are in place to recognise when the performance of this agent is not as expected, and thus, the RL application can no longer operate it
Level A	The RL application must be trained with real data of the agent that the application is to be deployed in. Safeguards are in place to recognise when the performance of the agent is not as expected, limiting the control by the RL application

unknown. We consider that this verification can be done through the following analyses:

- State to action mappings: the action output per state should be examined for an expected common range of states in the deployment environment. It must be demonstrated that no illegal, i.e. not allowed, action is taken. Additionally, it must be clearly defined when the RL application chooses to move to a temporary 'bad' state.
- Analysis of action sequences: from the state to action mappings, the expected sequence of actions over certain periods of time can be determined. Exemplary cases or normal operations shall be declared, in order to show the terminal state to which the RL application moves the operational environment is safe and desired.

Table 6 shows the advised certification compliance rules regarding the action output by the RL application.

Table 6 Proposed requirements regarding the action definition, per certification level.

Requirements - Action Definition	
Level D	The applicant presents the state-to-action mapping of the most common states, defining that illegal actions are not performed
Level C	State to action mapping and action sequence analysis of the most common states show that the agent does not perform illegal actions for the expected uses cases
Level B	State to action mapping and action sequence analysis of the most common states show that the RL agent does not perform illegal actions or move the operational environment to a temporary 'bad' state
Level A	State to action mapping and action sequence analysis show that the RL agent does not perform illegal actions or move the operational environment to a temporary 'bad' state. Safeguards are in place to recognise and protect against undesirable actions by the RL agent. The contingency plans for when such actions are detected should be clearly defined

F. Hazard and Risk Modelling

On top of the safeguards already defined in the previous sections, the maximum risk of deploying the RL application must be identified. It must be clear to what extent the application can negatively impact the deployment environment and what the likelihood of hazards occurring is. The applicant must declare the following:

- Hazard modelling and rare event simulation: before the deployment of the RL application, agent-based simulation tests must be performed to test potential hazards and rare events. The applicant must declare to what extent potential hazards, their impact and probability, are understood and tested.
- For safety-critical applications, risk assessment models for detecting future possible hazards must be developed. These shall include the detection of future dangerous states following the decisions of the RL application, as well as of rare states/events in which the application was not tested.

Table 7 shows the advised certification compliance rules regarding hazard and risk modelling of the RL model.

Table 7 Proposed requirements regarding hazard and risk modelling, per classification level.

Requirements - Hazard/Risk Modelling	
Level D	The applicant declares that the RL application is robust and has been tested to cover a fair number of cases
Level C	Extensive robustness tests are performed, covering the expected range of operations
Level B	Extensive robustness tests are performed, covering the expected range of operations. Safeguards are in place to protect against rare events
Level A	Extensive robustness testing covering all ranges of operation and worst-case scenarios. Safeguards are in place to predict future hazard situations and any state or action outside of the tested range of operations

G. Update of the RL Model/Renewal of the Certification

Changes to the RL application may require a renewal of the certification of the RL application. In decreasing order of magnitude, modifications to RL applications may contemplate:

- A core modification of the functioning of the RL application, e.g., changes to the state/action arrays, usage of a new RL algorithm, and modification of the hyperparameters of the application.
- Transfer learning, i.e., transferring knowledge from external expertise to improve the efficiency and the generalization abilities of the RL agent.
- Re-training of the application, e.g., update of the weights of the neural network nodes, leading to new actions to be performed per known states.
- Extended training of the application for new situations, i.e., the behaviour of the RL application is extended for a wider range of operations but the previous behaviour is not altered.

The following certification rules in Table 8 are advised in case of an update of the RL application.

Table 8 Proposed requirements regarding update/certification renewal of the RL model, per certification level.

Requirements - Update/Renewal of the RL Model	
Level D	The applicant declares that the update of the model will not hinder any existing functionality
Level C	Extensive robustness tests are performed, covering the expected range of operations
Level B	Re-training of the RL model leads to a new process of certification. Extended training for new situations can be certified as standalone only if demonstrated that they do not alter existing functionalities of the certified RL application
Level A	Any change to the RL model leads to a new process of certification

H. Documentation Requirements

Part of the advised necessary documents for certification compliance of RL applications is expected to match the current software certification guidelines [49] (e.g., software quality assurance plan, software design document, software requirements standard, software development plan, software verification plan, source code, test cases and procedures, verification results, problem reports). However, there are specific components unique to RL applications which must also be declared during a certification process:

- The files representing the policy of the RL application. Depending on the RL model, this may include several, or all, of the following elements: actor model, critic model, data existent in the replay buffer, the weight of all the nodes of the neural network layers, the composition of the hidden neural networks, activation functions, hyperparameters.
- Software, code, testing, and verification analysis of the simulated training conditions, including a comprehensive analysis of the simulated environment and simulated agents.
- Analysis of the necessary processing time for the RL application to output an action in response to a new state of the environment.

V. Discussion & Future Work

This paper presented an initial composition of a certification framework of RL applications in aviation, considering their technical implementation. Although this composition is meant as a stepping stone from which a more detailed framework can be developed, it already shows how this process must include:

- An analysis of the technical properties of the chosen RL algorithm. Certain algorithms entail a higher complexity for explainability and predictability of actions. Their usage must thus be justified in reference to the final deployment environment.
- An analysis of the characteristics during training of the model, the simulated environment, agents, actions, and training data should be delivered. It must be proven that these fairly represent the deployment environment to a level aligned with the criticality and autonomy of the RL application.
- Verification of how the RL application reacts against edge cases (i.e. states outside of the expected range of operations and/or not present in the training data). It should be clear whether a monitoring mechanism should be in place to protect against situations unknown to the RL application. Specifically, for safety-critical applications, RL should be combined with rare event simulation techniques.
- Declaration of uncertainties and perturbations of the deployment environment. These may include the reaction time of the deployment agent to start performing the action defined by the RL application, the potential for the deployment environment to evolve to unexpected states, and complex multi-agent interactions not fully represented in the simulated environment.

The authors aim to further refine this framework, hopefully in direct cooperation with regulatory entities. The most immediate steps shall focus on applying this framework to RL applications in aviation in the final stage of development. Such will allow us to identify (1) difficulties in compliance with the requirements defined in this paper, and (2) additional technical properties of RL methods that warrant inclusion in the framework. Finally, in a more advanced state, one of the objectives of this framework shall be to provide guidance on how to construct a simulation environment correctly to test RL applications and which benchmarks should be applied. Advanced simulation techniques should be developed to expose the application to a diverse set of scenarios and edge cases to validate its behaviour and identify use cases where its decision cannot be trusted.

VI. Conclusion

As of the date of this paper, regulatory entities have not yet established specific guidelines for the certification processes for RL applications in the aviation industry. However, a certification framework is essential to improve the public's trust in the safety and reliability of these systems. Additionally, the potential of RL applications in aviation can only be correctly identified once they are deployed in their intended operational environment. Moreover, a complete understanding of how the training environments differ from the operational environments and the consequences of this disparity may only be evaluated when the bridge between research and application is crossed. This delay in creating regulatory guidelines is attributed to the necessity of developing a certification framework tailored to the technical characteristics of RL implementations. We believe that this paper has shown a viable path for the definition of this framework.

The authors consider that the greater challenges with RL applications lay in their explainability, predictability, and verification that the training environment is a fair representation of the final deployment environment. The proposed framework defines that the certification applicant must present evidence justifying the RL algorithm employed, the fairness of the training environment in representing the final deployment environment, and the fairness of the simulated agents and their range of actions compared to the final deployment agent. Additionally, proper hazard and risk modelling representing edge cases and states unknown to the RL application is necessary for safety-critical applications. Finally, in safety-critical and autonomous RL applications, monitoring safeguards must be in place against undesirable behaviour by the application that may put agents or the environment at risk.

The authors intend to expand upon this work through direct collaboration with regulatory entities and researchers working on RL applications for air transport operations. Additionally, given the emphasis on the validation of the training environment, a primary objective is to enhance the certification guidelines to include benchmarks for these environments to properly identify the limitations of RL applications when transposed to deployment environments. Finally, future work shall delve into standardisation frameworks from other domains where RL has already been actively employed. Harmonizing standards across diverse research domains is expected to accelerate the development of the aviation framework.

References

[1] Li, Y., "Deep Reinforcement Learning: An Overview," *ArXiv*, Vol. abs/1701.07274, 2018.

[2] Razzaghi, P., Tabrizian, A., Guo, W., Chen, S., Taye, A., Thompson, E., Bregeon, A., Baheri, A., and Wei, P., "A Survey on Reinforcement Learning in Aviation Applications," , 2022.

[3] Degas, A., Islam, M. R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M. U., Begum, S., Rahman, M. A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., and Aricó, P., "A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory," *Applied Sciences*, Vol. 12, No. 3, 2022. https://doi.org/10.3390/app12031295.

[4] Blom, H., Everdij, M., and Bouarfa, S., *Complexity science in Air Traffic Management*, Addison-Wesley Professional, 2016.

[5] Ribeiro, M., Ellerbroek, J., and Hoekstra, J., "Distributed Conflict Resolution at High Traffic Densities with Reinforcement Learning," *Aerospace*, Vol. 9, No. 9, 2022. https://doi.org/10.3390/aerospace9090472.

[6] Brittain, M. W., Yang, X., and Wei, P., "Autonomous Separation Assurance with Deep Multi-Agent Reinforcement Learning," *Journal of Aerospace Information Systems*, Vol. 18, No. 12, 2021, pp. 890–905. https://doi.org/10.2514/1.I010973.

[7] Xu, D., Hui, Z., Liu, Y., and Chen, G., "Morphing control of a new bionic morphing UAV with deep reinforcement learning," *Aerospace Science and Technology*, Vol. 92, 2019, pp. 232–243. https://doi.org/https://doi.org/10.1016/j.ast.2019.05.058.

[8] Heyer, S., Kroezen, D., and Kampen, E.-J. V., "Online Adaptive Incremental Reinforcement Learning Flight Control for a CS-25 Class Aircraft," *AIAA Scitech 2020 Forum*, 2020. https://doi.org/10.2514/6.2020-1844.

[9] Tseremoglou, I., and Santos, B., "Condition-Based Maintenance Scheduling of an Aircraft Fleet Under Partial Observability: a Deep Reinforcement Learning Approach," *Reliability Engineering & System Safety*, Vol. 241, 2024. https://doi.org/10.1016/j.ress.2022.108908.

[10] Lee, J., and Mitici, M., "Deep reinforcement learning for predictive aircraft maintenance using probabilistic Remaining-Useful-Life prognostics," *Reliability Engineering & System Safety*, Vol. 230, 2022. https://doi.org/10.1016/j.ress.2022.108908.

[11] Dalmau, R., Ballerini, F., Naessens, H., Belkoura, S., and Wangnick, S., "An explainable machine learning approach to improve take-off time predictions," *Journal of Air Transport Management*, Vol. 95, 2021, p. 102090. https://doi.org/https://doi.org/10.1016/j.jairtraman.2021.102090.

[12] Memarzadeh, M., Akbari Asanjan, A., and Matthews, B., "Robust and Explainable Semi-Supervised Deep Learning Model for Anomaly Detection in Aviation," *Aerospace*, Vol. 9, No. 8, 2022. https://doi.org/10.3390/aerospace9080437.

[13] Parvez Farazi, N., Zou, B., Ahamed, T., and Barua, L., "Deep reinforcement learning in transportation research: A review," *Transportation Research Interdisciplinary Perspectives*, Vol. 11, 2021, p. 100425. https://doi.org/https://doi.org/10.1016/j.trip.2021.100425.

[14] Dmitriev, K., Schumann, J., and Holzapfel, F., "Toward Certification of Machine-Learning Systems for Low Criticality Airborne Applications," *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, 2021, pp. 1–7. https://doi.org/10.1109/DASC52595.2021.9594467.

[15] EASA, "EASA Artificial Intelligence Roadmap 2.0," https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-20-published, 2023.

[16] EASA, "First usable guidance for Level 1 & 2 machine learning applications," https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-concept-paper-proposed-issue-2-open, 2023.

[17] Balakrishna, P., Ganesan, R., and Sherry, L., "Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures," *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 6, 2010, pp. 950–962. https://doi.org/https://doi.org/10.1016/j.trc.2010.03.003, special issue on Transportation Simulation Advances in Air Transportation Research.

[18] Xiang, Z., Sun, H., and Zhang, J., "Application of Improved Q-Learning Algorithm in Dynamic Path Planning for Aircraft at Airports," *IEEE Access*, Vol. 11, 2023, pp. 107892–107905. https://doi.org/10.1109/ACCESS.2023.3321196.

[19] Lee, J., and Mitici, M., "Deep reinforcement learning for predictive aircraft maintenance using probabilistic Remaining-Useful-Life prognostics," *Reliability Engineering & System Safety*, Vol. 230, 2023, p. 108908. https://doi.org/https://doi.org/10.1016/j.ress.2022.108908.

[20] Dangut, M. D., Jennions, I. K., King, S., and Skaf, Z., "Application of deep reinforcement learning for extremely rare failure prediction in aircraft maintenance," *Mechanical Systems and Signal Processing*, Vol. 171, 2022, p. 108873. https://doi.org/https://doi.org/10.1016/j.ymssp.2022.108873.

[21] Brittain, M., and Wei, P., "Autonomous Air Traffic Controller: A Deep Multi-Agent Reinforcement Learning Approach," , 2019.

[22] Clarke, S. G., and Hwang, I., "Deep Reinforcement Learning Control for Aerobatic Maneuvering of Agile Fixed-Wing Aircraft," *AIAA Scitech 2020 Forum*, 2020. https://doi.org/10.2514/6.2020-0136.

[23] Kenworthy, L., Nayak, S., Chin, C., and Balakrishnan, H., "NICE: Robust Scheduling through Reinforcement Learning-Guided Integer Programming," , 2022.

[24] Shang, Z., Mao, Z., Zhang, H., and Xu, M., "Collaborative Path Planning of Multiple Carrier-based Aircraft Based on Multi-agent Reinforcement Learning," *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*, 2022, pp. 512–517. https://doi.org/10.1109/MDM55031.2022.00108.

[25] Xi, Z., Wu, D., Ni, W., and Ma, X., "Energy-Optimized Trajectory Planning for Solar-Powered Aircraft in a Wind Field Using Reinforcement Learning," *IEEE Access*, Vol. 10, 2022, pp. 87715–87732. https://doi.org/10.1109/ACCESS.2022.3199004.

[26] Yildiz, Y., Agogino, A., and Brat, G., "Predicting Pilot Behavior in Medium-Scale Scenarios Using Game Theory and Reinforcement Learning," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 4, 2014, pp. 1335–1343. https://doi.org/10.2514/1.G000176.

[27] Bewley, T., Lawry, J., and Richards, A., "Learning Interpretable Models of Aircraft Handling Behaviour by Reinforcement Learning from Human Feedback," , 2023.

[28] Alamdari, N., and Savard, G., "Deep reinforcement learning in seat inventory control problem: an action generation approach," *Journal of Revenue and Pricing Management*, Vol. 20, 2021. https://doi.org/10.1057/s41272-020-00275-x.

[29] Ikli, S., Mancel, C., Mongeau, M., Olive, X., and Rachelson, E., "The aircraft runway scheduling problem: A survey," *Computers & Operations Research*, Vol. 132, 2021, p. 105336. https://doi.org/https://doi.org/10.1016/j.cor.2021.105336.

[30] Limin, S., Due-Thinh, P., and Alam, S., "Multi-Agent Deep Reinforcement Learning for Mix-mode Runway Sequencing," *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 586–593. https://doi.org/10.1109/ITSC55140.2022.9922221.

[31] Azar, A. T., Koubaa, A., Ali Mohamed, N., Ibrahim, H. A., Ibrahim, Z. F., Kazim, M., Ammar, A., Benjdira, B., Khamis, A. M., Hameed, I. A., and Casalino, G., "Drone Deep Reinforcement Learning: A Review," *Electronics*, Vol. 10, No. 9, 2021. URL https://www.mdpi.com/2079-9292/10/9/999.

[32] EASA, "Concepts of Design Assurance for Neural Networks (CoDANN) II," https://www.easa.europa.eu/en/document-library/general-publications/concepts-design-assurance-neural-networks-codann-ii, 2021.

[33] FAA, " Certification Research Plan for Artificial Intelligence Applications (the Plan)," https://www.faa.gov/sites/faa.gov/files/NARP-03142023-National-Aviation-Research-Plan-2023-2027.pdf, 2023.

[34] NASA, "Challenges in the Verification of Reinforcement Learning Algorithms," https://ntrs.nasa.gov/api/citations/20170007190/downloads/20170007190.pdf, 2017.

[35] Baheri, A., Ren, H., Johnson, B., Razzaghi, P., and Wei, P., "A Verification Framework for Certifying Learning-Based Safety-Critical Aviation Systems," , 2022.

[36] Torens, C., Durak, U., and Dauer, J. C., "Guidelines and Regulatory Framework for Machine Learning in Aviation," *AIAA SCITECH 2022 Forum*, 2022. https://doi.org/10.2514/6.2022-1132.

[37] Delseny, H., Gabreau, C., Gauffriau, A., Beaudouin, B., Ponsolle, L., Alecu, L., Bonnin, H., Beltran, B., Duchel, D., Ginestet, J.-B., Hervieu, A., Martinez, G., Pasquet, S., Delmas, K., Pagetti, C., Gabriel, J.-M., Chapdelaine, C., Picard, S., Damour, M., Cappi, C., Gardès, L., Grancey, F. D., Jenn, E., Lefevre, B., Flandin, G., Gerchinovitz, S., Mamalet, F., and Albore, A., "White Paper Machine Learning in Certified Systems," , 2021.

[38] Gunning, D., Vorm, E., Wang, J. Y., and Turek, M., "DARPA's explainable AI (XAI) program: A retrospective," *Applied AI Letters*, Vol. 2, No. 4, 2021, p. e61. https://doi.org/https://doi.org/10.1002/ail2.61, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ail2.61.

[39] ICAO, "ICAO Safety Management Manual, Fourth Edition - 2018 (Doc 9859-AN/474)," , 2018.

[40] Sutton, R. S., and Barto, A. G., *Reinforcement learning: An introduction*, MIT press, 2018.

[41] Watkins, C. J. C. H., and Dayan, P., "Q-learning," *Machine Learning*, Vol. 8, No. 3, 1992, pp. 279–292. https://doi.org/10.1007/BF00992698.

[42] Rummery, G. A., and Niranjan, M., "On-Line Q-Learning Using Connectionist Systems," Tech. Rep. TR 166, Cambridge University Engineering Department, Cambridge, England, 1994.

[43] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M. A., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D., "Human-level control through deep reinforcement learning," *Nature*, Vol. 518, 2015, pp. 529–533.

[44] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D., "Continuous control with deep reinforcement learning," , 2019.

[45] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K., "Asynchronous Methods for Deep Reinforcement Learning," *Proceedings of The 33rd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 48, edited by M. F. Balcan and K. Q. Weinberger, PMLR, New York, New York, USA, 2016, pp. 1928–1937.

[46] Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S., "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," , 2018. Cite arxiv:1801.01290Comment: ICML 2018 Videos: sites.google.com/view/soft-actor-critic Code: github.com/haarnoja/sac.

[47] Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., Yang, Y., and Knoll, A., "A Review of Safe Reinforcement Learning: Methods, Theory and Applications," 2023.

[48] Wang, L., Yang, H., Lin, Y., Yin, S., and Wu, Y., "Explainable and Safe Reinforcement Learning for Autonomous Air Mobility," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 2022.

[49] RTCA, Incorporated, "DO-178C, Software Considerations in Airborne Systems and Equipment Certification," Tech. rep., 2011.