**3GPP TSG-SA5 Meeting #157 *S5-246254***

Hyderabad, India, 14 - 18 October 2024

**Source: ZTE Corporation**

**Title: Rel-19 pCR TR 28.858 Update the Description of Management of Reinforcement Learning Use Case**

**Document for: Approval**

**Agenda Item: 6.19.1**

# 1 Decision/action requested

***In this box give a very clear / short /concise statement of what is wanted.***

# 2 References

[1] 3GPP TR 28.858: " Study on Artificial Intelligence / Machine Learning (AI/ML) management Phase 2"

# 3 Rationale

The functionality of RL agent is captured in clause 5.1.7.1, where the RL agent can receive the current state of the environment, receive reward and selects action to be taken. However, Fig. 5.1.6.1 is unclear and may be misleading. Firstly, the definition for RL agent is missing but it is indicated as an ML Model in Fig. 5.1.6.1, which needs to be further clarified. Secondly, the reward setting mechanism is unclear. The environment cannot provide the reward to RL agent directly but the figure implies such interactions.

This contribution proposes to refine the description of RL agent and remove the Fig. 5.1.6.1.

# 4 Detailed proposal

***Start of First change***

### 5.1.7 Management of Reinforcement Learning

#### 5.1.7.1 Description

Reinforcement Learning (RL) is a type of machine learning where an agent learns to make decisions by taking actions in an environment to achieve a goal. The agent learns from the consequences of its actions, rather than from being explicitly taught and it selects its actions based on its past experiences (exploitation) and also by new choices (exploration), which is essentially a trial-and-error approach. In RL process, an agent interacts with the environment in discrete time steps. At each time step, the RL agent receives the current state of the environment and selects actions to be taken. The environment provides the new state in turn and the RL agent will receive the reward based on the state of the environment. For example, RL is applied for coverage problem analysis use case, the RL agent actions are used to optimize the coverage problem and the RL environment can be the simulation environment. The actions can be changes to the values of network adjustable parameters (e.g. change the transmission power of the NR sector carrier, see TS 28.104 [3]). The states can be the network PMs/KPIs like RSRP distribution etc. The reward can be the score of an RL performance metric to evaluate the PMs/KPIs. The goal of the agent is to learn a policy, which tells it what action to take under what circumstances, that maximizes the sum of rewards. The main advantage of this approach is in the ability to automatically adapt to the characteristics of the environment, making it suitable approach for handling dynamic environments such as mobile networks.

A screenshot of a computer

Description automatically generated

**Figure 5.1.6.1: Reinforcement Learning in Domain Management Function**

#### 5.1.7.2 Use cases

##### 5.1.7.2.1 Exploration in Reinforcement Learning

Reinforcement Learning (RL) has the ability to learn and adapt itself to dynamic environments and thus finds the near optimal solution to the problem. This makes the RL-based approaches very interesting for applications in the mobile networks. However, the potential negative impact to the mobile network caused by RL is still the main drawback. In particular, during the exploration step performing trials and learning from errors may have an impact on the operational network and may result in unsafe operations causing network performance degradations. Therefore, the exploration step in RL needs to be under a controlled environment that are not supposed to violate system performance requirements. If the RL agent behaves in an unexpected manner, there needs to be a set of fall-back actions in place, e.g., to switch from RL-based solution to non-RL-based solution.

For RL management, the MnS consumer can query the ML Training MnS producer to discover if RL is supported. When RL is supported, a consumer may want to provide a scope (e.g., geographical area, time window) that can aid the producer to select/create the environment when performing RL. In the event RL is supported, the consumer may also want to state their preference for environment type for RL during training i.e. simulated environment or real network.

NOTE: Support for both environment types can be considered optional in the RL training.

#### 5.1.7.3 Potential Requirements

**REQ-RL\_MGMT-01:** The ML training MnS producer should have a capability allowing an authorized MnS consumer to query if RL training is supported.

**REQ-RL\_MGMT-02:** The ML training MnS producer should have a capability allowing an authorized MnS consumer to specify the preferred RL environment type.

**REQ-RL\_MGMT-03:** The ML training MnS producer should have a capability to allow an authorized MnS consumer to specify the preferred RL environment scope

***End of First change***