**3GPP TSG-SA5 Meeting #157 *S5-245461***

Hyderabad, India, 14 - 18 October 2024

**Source: ZTE Corporation, Huawei, Nokia**

**Title: Rel-19 pCR TR 28.858 Add New Requirements and Possible Solution to the 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

When the RL process is performed in real network, the network performance degradations of unexpected actions selected by the RL agent should be considered and evaluated. To minimize the negative impact of RL process in the real network, AIML MnS Consumer should be allowed to specify fall back conditions for exploration in reinforcement learning.

This contribution proposes to add new requirement to RL use case.

# 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, an agent interacts with the environment in discrete time steps. At each time step, the agent receives the current state of the environment and selects an action. The environment, in turn, provides a reward and the new state. As illustrated in Figure 5.1.6.1, suppose 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 or a stable training configuration (higher exploration rate always results in a higher training efficiency with a more severe network performance degradation) that is 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, to fall back to last discrete time step, and to terminate the RL process.

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. When the real network is preferred by the MnS consumer, the consumer can provide network performance requirements (e.g. lower bound threshold, acceptable range, maximum performance deterioration Rate, etc.) of performing ML training of RL, so as to make the MnS producer adapt the training configurations to meet the network performance requirements.

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.

**REQ-RL\_MGMT-04:** The ML training MnS producer should have a capability to allow an authorized MnS consumer to specify RL configuration scope for exploration in reinforcement learning.

**REQ-RL\_MGMT-05:** The ML training MnS producer should have a capability allowing an authorized MnS consumer to provide network performance requirements of performing RL training.

#### 5.1.6.4 Possible solutions

##### 5.1.6.4.1 Possible solution #1

This solution proposes to enhance the existing MLTrainingRequest IOC and MLTrainingReport IOC to allow the RL consumer to specify RL requirement including environment selection and fall back condition.

**Enhancement on MLTrainingFunction IOC**:

Extend the existing IOC MLTrainingFunction with an attribute whose datatype is the supported learning technology including RL training, indicating the supported RL environment type (e.g. RL on simulation environments, RL on the real network) of the MnS producer.

**Enhancement on MLTrainingRequest IOC**: Introduce a new <<datatype>> “RLRequirement” to the MLTrainingRequest IOC, which may include following attributes:

* RLEnviromentType, represents preferred RL environment type. The allowed values may be “RealNetwork”, “SimulationEnviroment”:
* RLEnviromentScope, represents RL environment scope, which may be a geographical area, time window. .
* RLConfigurationScope, e.g., the range of actions that the agent is allowed to take.
* RLPerformanceRequirements, represents the attribute of the network performance requirements for online ML training, which indicates the tolerable network performance degradation (e.g. minimal/maximum performance value, maximum performance deterioration Rate). When the network performance is within the range, the RL training process can be continued. Otherwise, fall back actions can be determined by the producer.

### 5.1.7.5 Evaluation

The solution described in clause 5.1.6.4 proposes the addition of a new attribute to the MLTrainingFunction IOC to indicate the supported RL training and new attributes to the MLTrainingRequest IOC to enable the MnS consumer to indicate the preferred RL environment type and RL environment scope. Therefore, the solution described in clause 5.1.6.4 is a feasible solution to be developed further in the normative specifications.

***End of First change***