**3GPP TSG-SA5 Meeting #143-e S5-223384rev3**

**e-meeting, 9 - 17 May 2022**

**Source: Huawei, Nokia**

**Title: pCR TR 28.908 Add AI/ML training workflow and deployment scenarios for AI-ML management**

**Document for: Approval**

**Agenda Item: 6.5.5.1**

# 1 Decision/action requested

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

#  References

[1] SP-211443 New Study on AI/ ML management

[2] 3GPP TR 37.817 Study on enhancement for Data Collection for NR and EN-DC.

[3] 3GPP TR 28.813 Study on new aspects of Energy Efficiency (EE) for 5G.

[4] 3GPP TR 28.908 Study on Artificial Intelligence / Machine Learning (AI/ML) management

# 3 Rationale

The approved new SI [1] proposed to study the AI/ML management capabilities and management services to support/coordinate AI/ML in 5GS (3GPP management system, 5GC and NG-RAN). The AI/ML management workflow will helpful to illustrate the AI/ML management related capabilites and potential relationship.

In addition, one of the objective in SI [1] is investigation of deployment scenarios where the solutions are needed for AI/ML model training and each of the AI/ML model management capability. As discussed in RAN TR 38.817[2], OAM may need to be involved to support AI/ML capability in RAN. And as discussed in TR 28.813 [3], to provide centralized ES for RAN domain area, 3GPP management system need AI technology, so as to provide more efficient ES for RAN domain area while keeping basic KPIs stable for SLA assurance. Therefore, this contribution proposes to add AI/ML training deployment scenarios for AI-ML management.

This contribution proposes to add AI/ML management workflow and AI/ML training deployment scenarios.

# 4 Detailed proposal

It is proposed to add the following chapter in TR 28.908 [2].

|  |
| --- |
| **1st modified section** |

# X Concepts and overview

## X.Y AI/ML management workflow

Figure X.Y-1 illustrates the generic workflow for AI/ML management, which involves AI/ML training, AI/ML management, and AI/ML inference.



Figure X.Y-1: AI/ML management workflow

The AI/ML management related capabilities and potential relationship are described as following.

- **AI/ML training**: AI/ML training provides the capability that performs the AI/ML model training, validation, and testing. The AI/ML training includes initial model training and retraining. After the training is finished, the trained AI/ML model can be applied for inference.

- **AI/ML inference**: AI/ML inference provides the capability that generates the expected inference output based on the input inference data and a trained AI/ML model.

- **AI/ML management**: AI/ML management provides the capability of managing the entire process of using the AI/ML model, which includes:

* AI/ML training control: the AI/ML management can set the model retraining/updating period or trigger the model retraining/updating based on the performance evaluation results contained in the model performance feedback. For example, if the model performance decreases, the AI/ML performance management capability may trigger the AI/ML training to start retraining.
* AI/ML inference activation/deactivation/termination: the AI/ML management may select/activate/deactivate/terminate the AI/ML model and control the inference capability of the AI/ML Entity;

## X.Z AI/ML deployment scenarios with management

This clause provides the deployment scenarios for AI/ML management, which addresses the above AI/ML related capabilities including AI/ML training capability, AI/ML inference capability and AI/ML management capability.

As shown in the following figure, AI/ML cross domain training capability, AI/ML cross domain inference capability and AI/ML management capability resides on 3GPP (cross domain) management layer which manages the CN domain and RAN domain. AI/ML domain training capability and AI/ML domain inference capability resides on 3GPP domain management layer, where the AI/ML domain training capability and domain inference capability can be viewed as an entity to be managed. AI/ML RAN training capability, AI/ML RAN inference capability, AI/ML CN training capability and AI/ML CN inference capabilityresides on Network function layer.



Figure X.Z-1: AI/ML Deployment scenario

The above figure shows the overview of AI/ML related capabilites. Depending on different scenarios, different logical capabilties may be used.

Example Deployment scenario1: In this deployment scenario which provides intelligence in the 3GPP domain management layer, the AI/ML Management capability is located in 3GPP cross domain layer to provide management capability. The AI/ML domain training capability and AI/ML domain inference capability is located in 3GPP domain layer.

 

Figure X.Z-2: Example Deployment scenario1

Example Deployment scenario2: In this deployment scenario which supports the intelligence in RAN, the AI/ML Management capability is located in 3GPP cross domain layer to provide management capability. The AI/ML domain training capability is located in 3GPP domain layer to provide the AI/ML training capability to network function layer. The AI/ML RAN inference capability is located in Network function layer to provide inference capability.

 

Figure X.Z-3: Example Deployment scenario2

Notes: The AI/ML RAN inference Capability is not within the standard discussion scope of SA5, the split between the training capability and inference capability within the domain is up to the implementer.

|  |
| --- |
| **End of modified section** |