**3GPP TSG-SA5 Meeting #143-e S5-223385rev1**

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

**Source: Huawei**

**Title: pCR TR 28.908 Add Key Issue of AI/ML management for model deployment**

**Document for: Approval**

**Agenda Item: 6.5.5.4**

# 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 TS 28.104 Management and orchestration; Management Data Analytics (MDA)

[3] 3GPP TS 28.105 Management and orchestration; Artificial Intelligence / Machine Learning (AI/ML) management

[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). One of the AI/ML management capabilities is the deployment of AI/ML model. This contribution proposes to add a key issue of AI/ML management for model deployment.

# 4 Detailed proposal

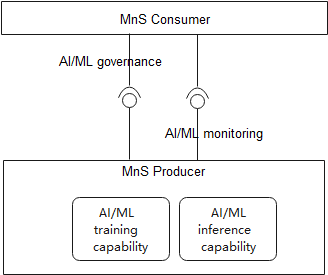
It is proposed to add the following chapter in TR 28.908 [4].

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

## X.Y Key Issue #XX: AI/ML management for model deployment

### X.Y.1 Description

The AI/ML training and inference capability can be viewed as an entity to be managed, which means the implementation of the internal capabilities and internal interactions between the steps could not be externally visible. However, some management capabilities (e.g. AI/ML governance and AI/ML monitoring) will exposed by the MnS producer, to enable the MnS consumer to manage the AI/ML training and inference. As following Figure X.Y.1-1.



**Figure X.Y.1-1: AI/ML model deployment**

AI/ML governance describes a set of capabilities to allow MnS consumer to govern AI/ML training capabilityand AI/ML inference capability, including:

- Lifecycle management of AI/ML training and inference, including deploy, update, activate/deactivate, terminate of the training and inference entity and the associated AI/ML model.

- Configure the period or trigger condition of model retraining/updating.

- Configure the reporting threshold of AI/ML monitoring.

AI/ML monitoring describes a set of capabilities to allow MnS consumer to monitoring the progress and result of AI/ML training and inference including:

- Monitor the AI/ML training and inference process status.

- Monitor the performance of the activated inference capabilityand the associated AI/ML model.

- Report the AI/ML model deployment status to MnS consumer.

This key issue studies the MnS of AI/ML governance and AI/ML monitoring to support AI/ML-enabled functions in 3GPP management system (MDA), NG-RAN (RAN intelligence) and 5GC (NWDAF).

### X.Y.2 Potential solutions

TBD.

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