**3GPP TSG-SA5 Meeting #154 *S5-241932d6***

Changsha, China, 15 - 19 April 2024

|  |  |
| --- | --- |
| *CR-Form-v12.1* |  |
| **CHANGE REQUEST** |  |
|  |  |
|  | **TS 28.105** | **CR** | **0129** | **rev** | **1** | **Current version:** | **18.3.0** |  |  |
|  |  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |  |
|  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network | **X** | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | Rel-18 CR TS 28.105 Update the description of ML model lifecycle management |
|  |  |
| ***Source to WG:*** | Huawei, Deutsche Telekom |
| ***Source to TSG:*** | S5 |
|  |  |
| ***Work item code:*** | AIML\_MGT |  | ***Date:*** | 2024-03-19 |
|  |  |  |  |  |
| ***Category:*** | F |  | ***Release:*** | Rel-18 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-15 (Release 15)Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)* |
|  |  |
| ***Reason for change:*** | Updated the description to make our AI/ML process more aligned with model lifecycle management. |
|  |  |
| ***Summary of change:*** | * Change the title of the 4a.0 and 6.1.
* Change the figure name and corresponding description.
* Remove phase related description.
 |
|  |  |
| ***Consequences if not approved:*** | The AI/ML terminology is unclear |
|  |  |
| ***Clauses affected:*** | 4a.0, 6.1 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

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

# 4a AI/ML management functionality and service framework

## 4a.0 ML model Lifecycle

AI/ML techniques are widely used in 5GS (including 5GC, NG-RAN, and management system), the generic AI/ML operational workflow in the lifecycle of an ML model is depicted in Figure 4a.0-1.



Figure 4a.0-1: ML model lifecycle

The ML model lifecyle includes training, emulation, deployment, and inference. These steps are briefly described below:

**- ML training:** training, including initial training and re-training of an ML model or a group of ML models. It also includes validation of the ML model to evaluate the performance when the ML model performs on the training data and validation data. If the validation result does not meet the expectations (e.g., the variance is not acceptable), the ML model needs to be re-trained.

**- ML testing:** testing of a validated ML model to evaluate the performance of the trained ML model when it performs on testing data. If the testing result meets the expectations, the ML model may proceed to the next step.

**- ML emulation:** running an ML model for inference in an emulation environment. The purpose is to evaluate the inference performance of the ML model in the emulation environment prior to applying it to the target network or system.

NOTE: The ML emulation is considered optional and can be skipped in the AI/ML operational workflow.

**- ML deployment:** ML deployment includes the ML model loading process (a.k.a. a sequence of atomic actions) which makes a trained ML model available for use at the target AI/ML inference function.

ML deployment step may not be needed in some cases, for example when the training function and inference function are co-located.

**- AI/ML inference:** performing inference using a trained ML model by the AI/ML inference function. The AI/ML inference may also trigger model re-training or update based on e.g., performance monitoring and evaluation.

|  |
| --- |
| **Next modified section** |

## 6.1 ML model Lifecycle management capabilities

Each operational step in the ML model Lifecycle (see clause 4a.0.1) is supported by one or more AI/ML management capabilities as depicted below.

**Management capabilities for ML training**

**- ML training management**: allowing the MnS consumer to request the ML training, consume and control the producer-initiated training, and manage the ML training/re-training process. The training management capability may include training performance management and setting a policy for the producer-initiated ML training.

ML training capability also includes validation to evaluate the performance of the ML model when performing on the validation data, and to identify the variance of the performance on the training and validation data. If the variance is not acceptable, the ML model would need to be re-trained before being made available for the next step in the operational workflow (e.g., ML model testing).

**Management capabilities for ML testing**

**- ML testing management**: allowing the MnS consumer to request the ML model testing, and to receive the testing results for a trained ML model. It may also include capabilities for selecting the specific performance metrics to be used or reported by the ML testing function. MnS consumer may also be allowed to trigger ML re-training based on the ML model testing performance requirements.

**Management capabilities for ML emulation:**

* **AI/ML inference emulation:** a capability allowing an MnS consumer to request an ML inference emulation for a specific ML model or entities (after the training, validation, and testing) to evaluate the inference performance in an emulation environment prior to applying it to the target network or system.

**Management capabilities for ML deployment:**

**- ML model loading management**: allowing the MnS consumer to trigger, control and/or monitor the ML model loading process.

**Management capabilities for AI/ML inference:**

**- AI/ML inference management:** allowing an MnS consumer to control the inference, i.e., activate/deactivate the inference function and/or ML models, configure the allowed ranges of the inference output parameters. The capabilities also allow the MnS consumer to monitor and evaluate the inference performance and when needed trigger an update of an ML model or an AI/ML inference function.

The use cases and corresponding requirements for AI/ML management capabilities are specified in the following clauses.

|  |
| --- |
| **End of modified sections** |