**3GPP TSG-SA5 Meeting #155 *S5-24331d18***

**Jeju, South Korea, 27 – 31 May 2024**

|  |
| --- |
| *CR-Form-v12.1* |
| **CHANGE REQUEST** |
|  |
|  | **28.105** | **CR** | **Input to Draft CR** | **rev** |  | **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 | **x** | Radio Access Network | **x** | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | Input to draft CR Rel-18 TS 28.105 further clarifications into emulation term instances |
|  |  |
| ***Source to WG:*** | NEC, Intel |
| ***Source to TSG:*** | S5 |
|  |  |
| ***Work item code:*** | AIML\_MGT |  | ***Date:*** | 2024-05-10 |
|  |  |  |  |  |
| ***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:*** | In the current specifications the emulation term is sometimes interchangeably expressed with the prefix ML or AI/ML. In some instances, it is coupled with the term “inference” and some other times it is not. |
|  |  |
| ***Summary of change:*** | Improve consistency and align the terms relating to the emulation while also considering the associated context.  |
|  |  |
| ***Consequences if not approved:*** | Unnclear, duplicated and/or mising term definitions may lead to misunderstanding and unnecessary speculations of the specifications potentially leading to implementation problems. |
|  |  |
| ***Clauses affected:*** | 4a.0, 6.10, 6.3, 7.3a.2 |
|  |  |
|  | **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 input to draft CR is using the latest baseline Rel-18 TS28.105 v18.3.0 with change marks.  |
|  |  |
| ***This CR's revision history:*** |  |

***1st change***

## 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 model 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 entity to evaluate the performance when the ML entity 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 associated with that entity needs to be re-trained.

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

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

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

**- ML model deployment:** ML deployment includes the ML model loading process (a.k.a. a sequence of atomic actions) to make 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 entity 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.

***2nd change***

## 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 model 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 entity 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 entity would need to be re-trained before being made available for the next step in the operational workflow (e.g., ML entity testing).

**Management capabilities for ML testing**

**- ML model testing management**: allowing the MnS consumer to request the ML entity testing, and to receive the testing results for a trained ML entity. 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 entity testing performance results.

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

* **AI/ML inference emulation:** a capability allowing an MnS consumer to request an AI/ML inference emulation for a specific ML entity 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 entity loading management**: allowing the MnS consumer to trigger, control and/or monitor the ML entity 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 entity/entities, 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 entity or an AI/ML inference function.

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

***3rd change***

6.3 AI/ML inference emulation

6.3.1 Description

Before the ML entity is applied in the production network, the MnS inference consumer may want to receive results of inference in one or more environments that emulate (to different extents) the expected inference characteristics, in a process that may be termed as inference emulation. The inference emulation enables this.

6.3.2 Use cases

6.3.2.1 AI/ML inference emulation

After the ML entity is validated and tested during development, the MnS consumer may wish to receive information from an inference emulation process that indicates if the ML entity or the associated ML inference function is working correctly under certain runtime context.

The management system should have the capabilities enabling an MnS consumer:

- request an inference emulation function to provide emulation reports; and

- to receive the results from running inference through an AI/ML inference emulation environment available at the emulation MnS producer.

6.3.3 Requirements for Managing AI/ML inference emulation

**Table 6.3.3-1**

| **Requirement label** | **Description** | **Related use case(s)** |
| --- | --- | --- |
| **REQ-AI/ML\_EMUL-1:** | The MnS producer for AI/ML inference emulation should have a capability enabling an authorized MnS consumer to receive reporting about the ML inference emulation.  | AI/ML Inference emulation (clause 6.3.2.1) |
| **REQ-AI/ML\_EMUL-2:** | The MnS producer for AI/ML inference emulation should have a capability enabling an authorized MnS consumer to request an inference emulation function to provide inference emulation reports on an ML entity or inference Function. | AI/ML Inference emulation (clause 6.3.2.1) |

***4th change***

7.3a.2 Information model definitions for AI/ML inference emulation

7.3a.2.1 Class diagram

7.3a.2.1.1 Relationships



**Figure 7.3a.2.1.1-1: NRM fragment for AI/ML inference emulation Control**

7.3a.2.1.2 Inheritance



Figure 7.3a.2.1.2-1: AI/ML inference emulation Inheritance Relations

***End of changes***