**3GPP TSG-SA2 Meeting #155 *S2-2303467***

**Athens, Greece, Feb 20th – 24th,2023 (revision of S2-2302469)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.2* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | **23.288** | **CR** | **0557** | **rev** | **4** | **Current version:** | **18.0.0** |  |
|  | | | | | | | | |
| *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 |  | Core Network | **X** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | Monitoring of accuracy of ML models | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Ericsson, [CATT], Vivo | | | | | | | | | |
| ***Source to TSG:*** | SA2 | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | eNA\_Ph3 | | | | |  | ***Date:*** | | | 2023-01-31 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | B |  | | | | | ***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 can be 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-16 (Release 16) Rel-17 (Release 17) Rel-18 (Release 18) Rel-19 (Release 19)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | Adding support for monitoring of accuracy of ML models, as agreed in TR 23.700-81  SA2#155:  there is an EN:  Editor's note: It is FFS whether the NWDAF containing AnLF registers to the NWDAF containing MTLF for monitoring analytics accuracy of the ML model using the procedure in clause 6.2E.3.2 or reusing the ML Model Provisioning procedures (e.g. including capability of monitoring analytics accuracy information of the ML model in the Nnwdaf\_MLModelProvision\_Subscribe request).  In the conclusion of KI#1 in TR 23.700-81, we have concluded that:  - MTLF subscribes to AnLF, that is registered in MTLF with its accuracy monitoring for a model provided by that MTLF, for getting notifications of the accuracy degradation of the analytics generated by the model, where the AnLF determines accuracy information based on any of the following:  - Comparing predictions and its corresponding ground truth data.  The motivation of AnLF registers in MTLF of model monitoring is MTLF does not have information about if the AnLF will use the model and when it will use the model and monitor the analytics accuracy. It is beneficial that the AnLF can inform MTLF when it starts or stops using / monitoring the accuracy of the model by using (de)registraion operation. Only extending model provisioning request cannot achive the purpose, since AnLF cannot determine whether and when it will use the model before the model is downloaded, which can result in half open subscriptions from MTLF to AnLF (i.e., subscription that will never receive notifications) Furthermore, extending ML Model provisioning service for this purpose will impact the semantics of the service and is not concluded in the study phase.  Regarding another EN:  Editor's Note: It is FFS how the NWDAF containing MTLF determine the ML model degradation based on the multiple analytics accuracy reports received from multiple AnLFs.  In CR - 0673, the following concept was introduced,  “*- [OPTIONAL] Minimal number of analytics output occurrences: determines the minimal number of analytics output provided by NWDAF that have to be considered in the determination of the accuracy information.*”  Inspired by this idea, if the number of inferences is provided by multiple NWDAFs containing AnLF, the NWDAF containing MTLF can use them to calculate a global accuracy, such as:    The NWDAF containing MTLF can use this global accuracy to determine if the ML model is degraded.  Given that in clause 5C.1, it is explained how MTLF determine model degradation, the analytics accuracy feedback from AnLF is one of the factors to consider, we can simplify the above EN in clause  How to determine if a model is degradated is MTLF internal logic, we can add an NOTE to explain that. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | ML Model information is enhanced with ML Model identifier and degraded indication.  New procedures for monitoring the accuracy of a model  SA2#155:  remove the EN:  Editor's note: It is FFS whether the NWDAF containing AnLF registers to the NWDAF containing MTLF for monitoring analytics accuracy of the ML model using the procedure in clause 6.2E.3.2 or reusing the ML Model Provisioning procedures (e.g. including capability of monitoring analytics accuracy information of the ML model in the Nnwdaf\_MLModelProvision\_Subscribe request).  Adding NOTE: The actual mechanism for the NWDAF containing MTLF for determining whether ML model degradation has taken place, based on the multiple analytics accuracy reports received from NWDAF containing AnLF, is an internal procedure of the NWDAF containing MTLF, e.g., the NWDAF containing MTLF calculate a global accuracy based on the analytics accuracy information and the number of inferences received from multiple NWDAFs containing AnLF. The details are not subject to standardization. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | No support for Monitoring the accuracy of an ML Model | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 6.2A.2, New 6.2E.3 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **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:*** | |  | | | | | | | | |

\*\*\*\*\*\*\*\*\*\* Start of Changes \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

### 6.2A.2 Contents of ML Model Provisioning

The consumers of the ML model provisioning services (i.e. an NWDAF containing AnLF) as described in clause 7.5 and clause 7.6 may provide the input parameters as listed below:

- Information of the analytics for which the requested ML model is to be used, including:

- A list of Analytics ID(s): identifies the analytics for which the ML model is used.

- [OPTIONAL] Use case context: indicates the context of use of the analytics to select the most relevant ML model ML model.

NOTE 1: The NWDAF containing MTLF can use the parameter "Use case context" to select the most relevant ML model, when several ML models are available for the requested Analytics ID(s). The values of this parameter are not standardized.

- [OPTIONAL] ML Model Interoperability Information. This is vendor-specific information that conveys, e.g., requested model file format, model execution environment, etc. The encoding, format, and value of ML Model Interoperable Information is not specified since it is vendor specific information, and is agreed between vendors, if necessary for sharing purposes.

- [OPTIONAL] ML Model Filter Information: enables to select which ML model for the analytics is requested, e.g. S-NSSAI, Area of Interest. Parameter types in the ML Model Filter Information are the same as parameter types in the Analytics Filter Information which are defined in procedures.

- [OPTIONAL] Target of ML Model Reporting: indicates the object(s) for which ML model is requested, e.g. specific UEs, a group of UE(s) or any UE (i.e. all UEs).

- ML Model Reporting Information with the following parameters:

- (Only for Nnwdaf\_MLModelProvision\_Subscribe) ML Model Reporting Information Parameters as per Event Reporting Information Parameter defined in Table 4.15.1-1, TS 23.502 [3].

- [OPTIONAL] ML Model Target Period: indicates time interval [start, end] for which ML model for the Analytics is requested. The time interval is expressed with actual start time and actual end time (e.g. via UTC time).

- A Notification Target Address (+ Notification Correlation ID) as defined in clause 4.15.1 of TS 23.502 [3], allowing to correlate notifications received from the NWDAF containing MTLF with this subscription.

- [OPTIONAL] Indication of supporting multiple ML models.

- [OPTIONAL] accuracy level of Interest.

Editor's note: It is FFS if additional parameters are needed for multiple model provisioning.

The NWDAF containing MTLF provides to the consumer of the ML model provisioning service operations as described in clause 7.5 and 7.6, the output information as listed below:

- (Only for Nnwdaf\_MLModelProvision\_Notify) The Notification Correlation Information.

- ML Model Information, which includes:

- the ML model file address (e.g. URL or FQDN) for the Analytics ID(s), when multiple ML models is not supported; or.

- a set of pair of unique ML Model identifier and the ML model file address (e.g. URL or FQDN) for the Analytics ID(s), if multiple ML models is supported.

Editor's note: It is FFS how to convey degradation information for an ML model.

- [OPTIONAL] Validity period: indicates time period when the provided ML Model Information applies.

- [OPTIONAL] Spatial validity: indicates Area where the provided ML Model Information applies.

NOTE 2: Spatial validity and Validity period are determined by MTLF internal logic and it is a subset of AoI if provided in ML Model Filter Information and of ML Model Target Period, respectively.

\*\*\*\*\*\*\*\*\*\* Next Change \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

### 6.2E.3 Procedure for AnLF-assisted MTLF ML Models Accuracy Monitoring

#### 6.2E.3.1 General

The procedures described in this clause enable the following functionality:

* An NWDAF containing AnLF may register with an NWDAF containing MTLF when it starts using an ML model and monitoring the accuracy of analytics generated by that ML Model for a given Analytics ID. It is assumed that the NWDAF containing AnLF obtained the ML model in a previous interaction with the NWDAF containing MTLF, e.g., using the Nnwdaf\_MLModelInfo\_Request or Nnwdaf\_MLModelProvision services. This registration enables the NWDAF containing MTLF to become aware of NWDAF containing AnLF that are using a given ML model for certain Analytics ID and that the NWDAF containing AnLF supports the capability of monitoring the accuracy of the corresponding analytics.
* An NWDAF containing MTLF may subscribe to an NWDAF containing AnLF where an existing Nnwdaf\_MLModelProvision service is established for receiving notifications of the accuracy of analytics generated by a given ML model for a certain Analytics ID. NWDAF containing AnLF can generate the accuracy information in many ways: e.g. comparing predictions of ML model and its corresponding ground truth data, comparing changes in internal configuration for the analytics ID generation, previous existent records of analytics accuracy information etc.

Editor's note: The service names in the following procedures are descriptive and to be updated.

#### 6.2E.3.2 Procedures for registering the monitoring of the analytics accuracy of an ML Model

When an NWDAF containing AnLF starts making use of an ML model and it has the ability to monitor the analytics accuracying the accuracy of an ML model, it registers with the NWDAF containing MTLF, that is responsible for training/updating this ML model.

When the NWDAF containing AnLF is no longer using the ML model or monitoring the accuracy of the analytics generated by that ML model for the Analytics ID, it de-registers it with the responsible NWDAF containing MTLF.

Figure 6.2E.3.2-1 illustrates the procedure by which an NWDAF containing AnLF registers with an NWDAF containing MTLF that it is starting to make use and monitor the analytics accuracy of an ML model. A new Nnwdaf\_MLModelMonitor\_Register service operation is used for that purpose.



Figure 6.2E.3.2-1: Procedure for ML Model monitoring registration

An NWDAF containing AnLF may start monitoring the accuracy of an ML model based on localy policy or request from its service consumer.

1-2. The NWDAF containing AnLF sends an Nnwdaf\_MLModelMonitor\_Register request to an NWDAF containing MTLF (NWDAF containing AnLF NF ID, unique identifier of the ML model, optionally: subscription endpoint of the Nnwdaf\_MLModelMonitor\_Subscribe service operation at the NWDAF containing AnLF). The NWDAF containing MTLF is now aware of the NF ID of the NWDAF containing AnLF that is monitoring the accuracy of that ML model.

When the NWDAF containing AnLF does no longer monitor the analytics accuracy of the ML model, it sends an Nnwdaf\_MLModelMontior\_Deregister service operation. The NWDAF containing MTLF deletes the associated data.

#### 6.2E.3.3 Procedures for monitoring the analytics accuracy of an ML model

An NWDAF containing MTLF, due to the registration of monitoring of the analytics accuracy of an ML model received from NWDAF containing AnLF and local policies, subscribes to the NWDAF containing AnLF for receiving notifications of the accuracy of the ML Model. The NWDAF containing MTLF may get the Subscription endpoint address of the NWDAF containing AnLF from the information received in a previous registration or through a service discovery procedure at the NRF.

Figure 6.2E.3.3-1 illustrates the procedure for monitoring the analytics accuracy of an ML model. Nnwdaf\_MLModelMonitor\_Subscribe and Nnwdaf\_MLModelMonitor\_Notify service operations are used for that purpose. A service consumer, i.e., an NWDAF containing MTLF, subscribes at a service producer, i.e., an NWDAF containing AnLF, to be notified when the analytics accuracy of the previously provisioned ML model is not sufficient.



Figure 6.2E.3.3-1: Procedure for monitoring the analytics accuracy of an ML model

0. Upon the reception of an Nnwdaf\_MLModelMonitor\_Register request and based on local policy, the NWDAF containing MTLF determines to subscribe to the analytics accuracy monitoring for the ML model.

1. The NWDAF containing MTLF sends an Nnwdaf\_MLModelMonitor\_Subscribe request (Analytics ID(s), unique identifier(s) of the ML model(s) to be monitored, accuracy metrics to be monitored, optionally Reporting Threshold(s) or Reporting Period) to an NWDAF containing AnLF subscription endpoint.

2. The NWDAF containing AnLF sends a response to the NWDAF containing MTLF.

3. The NWDAF containing AnLF starts monitoring the analytics accuracy of the ML model(s), if it not started yet.

NOTE 1: The NWDAF containing AnLF can monitor the analytics accuracy in many ways: e.g. comparing predictions of ML model and its corresponding ground truth data, comparing changes in internal configuration for the analytics ID generation, previous existent records of analytics accuracy information etc.

4-5. When the analytics accuracy of the ML model is insufficient i.e., deviation of the output analytics using the trained ML model from ground truth data (which are collected from Data Producer NF corresponding to analytic ID requested at the time which the prediction refers to) is greater than the Reporting Threshold(s) (which are locally configured or received in the Subscribe request), or the Reporting Period indicated in the Subscribe request is reached, the NWDAF containing AnLF sends an Nnwdaf\_MLModelMonitor\_Notify request to the notification endpoint (e.g., the NWDAF containing MTLF). The Notify request includes the analytics accuracy information of the ML model (e.g., a Deviation value which indicates the deviation of the predictions generated using the ML model(s) from the data, and the network data when the deviation occurs (which can be used by the NWDAF containing MTLF for possible ML model retraining), and the number of inferences were performed), and optionally an indication that the analytics accuracy of the ML model does not meet the requirement of accuracy for the ML model.

6. The NWDAF containing MTLF sends a response.

7. When the NWDAF containing MTLF has received the multiple analytics accuracy information, from one or more NWDAFs containing AnLF, it may consider that the ML model is degraded/to be updated (i.e., enough number analytics accuracy information received from one or more NWDAFs containing AnLF, indicating insufficient analytics accuracy).

NOTE 2: The actual mechanism for the NWDAF containing MTLF for determining the degradation of the ML model degradation is an internal procedure of the NWDAF containing MTLF, e.g., the NWDAF containing MTLF calculate a global accuracy based on the analytics accuracy information and the number of inferences received from multiple NWDAFs containing AnLF. 8. When an ML model is considered degraded / to be updated, the NWDAF containing MTLF re-trains the existing ML model or selects a new ML model. If the network data was not included in the Nnwdaf\_MLModelMonitor\_Notify request of step 5, the NWDAF containing MTLF may request data from the NWDAF containing AnLF, ADRF and/or other 5GS entities as specified in clause 6.2, and use the collected data for ML model retraining. The NWDAF containing MTLF notifies the NWDAF(s) containing AnLF with the updated trained ML Model Information by invoking Nnwdaf\_MLModelProvision\_Notify service operation, as described in clause 6.2A.

\*\*\*\*\*\*\*\*\*\* End of Changes \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*