**3GPP TSG-SA5 Meeting #143eS5-223395**

**09 - 17 May 2022, E-meeting**

**Source: Nokia**

**Title: pCR 28.105 Add requirements for pre-processed event data for ML training**

**Document for: Approval**

**Agenda Item: 6.6.5**

# 1 Decision/action requested

***The group is asked to discuss and agree on the proposal.***

# 2 References

[1] 3GPP TS 28.105-110 “Management and orchestration; AI/ML management”.

# 3 Rationale

The AI/ML training needs to support the capabilities for handling errors in input data or in the learned decisions. This pCR presents the corresponding requirements.

# 4 Detailed proposal

|  |
| --- |
| **Start of modifications** |

## 6.N Pre-processed event data for ML training

### 6.N.1 Description

In analytics solutions, Performance Measurements (PMs) and Fault Reports (FRs) from various network function are collected and analytics applied on the PMs and FRs to come up with statistical insights and predictions of events from the raw data. For most algorithms, the prediction accuracy depends upon the amount of relevant historical data, motivating the need to store ever more data, which correspondingly increases the storage and processing resource requirements. However, not all recorded data is useful as the derived events, e.g. captured through analytics processes, may have loss of information OR mis-information e.g., with respect to time of the event.

Relatedly for AI/ML algorithms, a large amount of data points does not necessarily add value, e.g., if most of it includes biased data which ends up getting discarded during the pre-processing stages. Instead, the AI/ML algorithms need to have information-rich events data that is condensed but with most of it useful for the required training. For example, one could train an interference optimization solution that learns the best way to combat interference by looking at counters of handover failures correlated with signal quality. However, for most of the time in the radio network, there will be no interference events but this cannot be determined if the events are not captured form the data. As such all the data must be kept and used for training. However, the data could also be mined for the interference event or signatures thereof. Then an equivalent interference management solution could be trained using the interference signatures or the signatures combined with only a small amount of raw data.

It is as such necessary to provide means to isolate and store the information rich events in the network, to ensure that minimizing storage and processing costs by discarding the unnecessary raw data does not compromise the ability to and still avails adequate historical information to adequately train AI/ML applications. In other words it is necessary for network functions to their management system to generate data on about the observed network events, e.g., based on the criteria set by the Operator, which events can then be stored to be used later to train AI/ML applications.

Network Function

AI/ML enabled Function as Network events Data Consumer

Network Function

Events processing

Events data

Management Function

Events processing

Network/system data

Network/system events data

 Fig 3. Exposing and storing network events data

### 6.N.2 Requirements

|  |  |  |
| --- | --- | --- |
| **Requirement label** | **Description** | **Related use case(s)** |
| **ML\_Error\_Req\_1** | The 3GPP management system shall enable an authorized consumer to request from the network data producer for network events corresponding to the data produced by that network data producer. | Pre-processed event data for ML training |
| **ML\_Error\_Req\_2** | The 3GPP management system shall enable A network data producer to generate network events in place of or alongside the network data that they produce | Pre-processed event data for ML training |
| **ML\_Error\_Req\_3** | The 3GPP management system shall enable a network events aggregator to take the events from different network entities and re-expose them in an aggregated way that eliminates duplications  | Pre-processed event data for ML training |

|  |
| --- |
| **End of modifications** |