**3GPP TSG-RAN3 Meeting #114-e R3-216192**

**E-meeting, 1 - 11 Nov 2021**

**Title:** TP for TR 37.817 on General Framework

**Source:** Deutsche Telekom

**Agenda item:** 18.2

**Document Type:** Other

# TP to TR 37.817

<<<< Start of changes >>>>

# 4 General Framework

*Editor Note: high level principles for RAN intelligence enabled by AI, the functional framework (e.g. the AI functionality and the input/output of the component for AI enabled optimization)*

4.1 High-level Principles

The following high-level principles should be applied for AI-enabled RAN intelligence:

* The detailed AI/ML algorithms and models for use cases are implementation specific and out of RAN3 scope.
* The study focuses on AI/ML functionality and corresponding types of inputs/outputs.
* The input/output and the location of the Model Training and Model Inference function should be studied case by case.
* The study focuses on the analysis of data needed at the Model Training function from Data Collection, while the aspects of how the Model Training function uses inputs to train a model are out of RAN3 scope.
* The study focuses on the analysis of data needed at the Model Inference function from Data Collection, while the aspects of how the Model Inference function uses inputs to derive outputs are out of RAN3 scope.
* Where AI/ML functionality resides within the current RAN architecture, depends on deployment and on the specific use cases.
* The Model Training and Model Inference functions should be able to request, if needed, specific information to be used to train or execute the AI/ML algorithm and to avoid reception of unnecessary information. The nature of such information depends on the use case and on the AI/ML algorithm.
* The Model Inference function should signal the outputs of the model only to nodes that have explicitly requested them (e.g. via subscription), or nodes that are subject to actions based on the output from model inference.
* An AI/ML model used in a Model Inference function has to be initially trained, validated and tested before deployment.
* NG-RAN is prioritized; EN-DC is included in the scope. FFS on whether MR-DC should be down-prioritized.
* A general framework and workflow for AI/ML optimization should be defined and captured in the TR. The generalized workflow should not prevent to “think beyond” the workflow if the use case requires so.
* User data privacy and anonymisation should be respected during AI/ML operation.

4.2 Functional Framework

*Editor Note: FFS on whether model performance evaluation / generating of model performance metrics is performed in Model Inference.*



Figure 4.2-1: Functional Framework for RAN Intelligence

This section introduces the common terminologies related to the functional framework for RAN intelligence illustrated in Figure 4.2-1. For the functions and data/information flows shown in the Figure 4.2-1, whether there is any standardization impact and what is the standardization impact are discussed in clause 5.

* Data Collection is a function that provides input data to Model Training and Model Inference functions. AI/ML algorithm specific data preparation (e.g., data pre-processing and cleaning, formatting, and transformation) is not carried out in the Data Collection function.   
  Examples of input data may include measurements from UEs or different network entities, feedback from Actor, output from an AI/ML model.
  + Training Data: Data needed as input for the AI/ML Model Training function.
  + Inference Data: Data needed as input for the AI/ML Model Inference function.
* Model Training is a function that performs the ML model training, validation, and testing which may generate model performance metrics as part of the model testing procedure. The Model Training function is also responsible for data preparation (e.g., data pre-processing and cleaning, formatting, and transformation) based on Training Data delivered by a Data Collection function, if required.
  + Model Deployment/Update: Used to initially deploy a trained, validated, and tested AI/ML model to the Model Inference function or to deliver an updated model to the Model Inference function. 
    - Note: Details of the Model Deployment/Update process as well as the use case specific AI/ML models transferred via this process are out of RAN3 Rel-17 study scope. The feasibility to single-vendor or multi-vendor environment has not been studied in RAN3 Rel-17 study.

* Model Inference is a function that provides AI/ML model inference output (e.g. predictions or decisions). It is FFS whether it provides model performance feedback to Model Training function. The Model Inference function is also responsible for data preparation (e.g. data pre-processing and cleaning, formatting, and transformation) based on Inference Data delivered by a Data Collection function, if required.
  + Output: The inference output of the AI/ML model produced by a Model Inference function.
    - Note: Details of inference output are use case specific.
  + (FFS) Model Performance Feedback: Applied if certain information derived from Model Inference function is suitable for improvement of the AI/ML model trained in Model Training function. Feedback from Actor or other network entities (via Data Collection function) may be needed at Model Inference function to create Model Performance Feedback.
    - Note: Details of the Model Performance Feedback process are out of RAN3 Rel-17 study scope.
* Actor is a function that receives the output from the Model Inference function and triggers or performs corresponding actions. The Actor may trigger actions directed to other entities or to itself.
  + Feedback: Information that may be needed to derive training or inference data or performance feedback.

<<<< End of changes>>>>