RP-233796

Deutsche Telekom
& T-Mobile USA

3GPP RAN#102 | Dec. 11th - 15th 2023
Edinburgh, Scotland, UK

Need for a consistent e2e AI/ML framework

Some observations

What has been done/agreed:

- SA WGs 1 - 5, CT WGs 3 – 4, and RAN WGs 1 - 3 have already performed SIs and/or WIs on the AI/ML topic.
- They addressed use cases from their own perspective (e.g. NWDAF usage in the CN, MDAS in OAM, AI/ML for NG-RAN and for air interface in the RAN domain), but a consolidated E2E framework, especially from a lifecycle management (LCM) perspective for AI/ML models, is still missing.
- AI/ML models are certainly implementation specific and therefore out of scope of 3GPP, but MNOs see the need to have a common LCM framework at least for those models that have impact on the network and E2E performance.

What is missing:

- A common E2E LCM framework for AI/ML models describing
  - how models are trained, validated, and tested,
  - how deployment of models to inference machines (as part of logical NW nodes or functions) is handled,
  - how models are updated, replaced and/or finally released,
  - how model inference output is ensured not to negatively impact the NW and E2E performance,
  - with all issues under the control of the MNO.
- As TSG SA is responsible for the E2E architecture, the common E2E LCM framework should be defined or described by related SA WGs (e.g., SA2 in collaboration with SA5) taking into account the status of existing and ongoing work in other SA, CT, and RAN WGs (continuous LS exchange would be appreciated).

Our urgent proposal for Rel-19

As TSG SA is responsible for the E2E architecture, the common E2E LCM framework should be defined or described by related SA WGs (e.g., SA2 in collaboration with SA5) taking into account the status of existing and ongoing work in other SA, CT, and RAN WGs (continuous LS exchange would be appreciated).

Thanks.