3GPP TSG RAN WG1 #118 R1- 2405962

Maastricht, NL, 19 Aug 2024 - 23 Aug 2024

**Agenda item:** **9.1.3.3**

**Source: Tejas Networks.**

**Title:** **Other aspects of AI/ML Model and Data**

**Document for: Discussion and Decision**

# Discussion

RAN #102 meeting approved the Rel-19 WI on AI/ML for NR Air Interface [1] based on that the AI/ML techniques to NR air interface has been studied in FS\_NR\_AIML\_Air [2].

|  |
| --- |
| **Objectives in RP-234039**  Study objectives with corresponding checkpoints in RAN#105 (Sept ’24):  […]   * Necessity and details of model Identification concept and procedure in the context of LCM [RAN2/RAN1] * CN/OAM/OTT collection of UE-sided model training data [RAN2/RAN1]:   + For the FS\_NR\_AIML\_Air study use cases, identify the corresponding contents of UE data collection   + Analyse the UE data collection mechanisms identified during the FS\_NR\_AIML\_Air (TR 38.843 section 7.2.1.3.2) study along with the implications and limitations of each of the methods * Model transfer/delivery [RAN2/RAN1]:   + Determine whether there is a need to consider standardised solutions for transferring/delivering AI/ML model(s) considering at least the solutions identified during the FS\_NR\_AIML\_Air study   […]  NOTE: offline training is assumed for the purpose of this project.  NOTE: the outcome of the study objectives should be captured in TR 38.843 for future reference.  NOTE: Coordination with SA/SA WGs of the ongoing study/work as it may relate to their required work. |

In this contribution we provide discussions on Model identification and model transfer/delivery based on previous RAN meetings and TR 38.843.

# Model identification

In RAN1 #116, #116bis and #117 meetings the following agreements related to model identification have been reached.

|  |
| --- |
| Agreement#116   * To facilitate the discussion, RAN1 studies the model identification type A with more details related to use cases. * To facilitate the discussion, RAN1 studies the following options as starting point for model identification type B with more details related to all use cases * MI-Option 1: Model identification with data collection related configuration(s) and/or indication(s) * MI-Option 2: Model identification with dataset transfer * MI-Option 3: Model identification in model transfer from NW to UE * FFS: The boundary of the options * Note: the names (MI-Opton1, MI-Option 2, MI-Option 3) are used only for discussion purpose * Note: other options are not precluded   **Observation**  The other options are proposed for model identification type B by companies during the discussion:   * MI-Option 4. Model identification via standardization of reference models. (for CSI compression) * MI-Option 5. Model identification via model monitoring   **Agreement #116**   * Regarding MI-Option 1 (Model identification with data collection related configuration(s) and/or indication(s)) of model identification type B, RAN1 further study the following aspects:   + Relationship between model ID and data collection related configuration(s) and/or indication(s)   + Information transmitted from NW to UE (if any)   + Information transmitted from UE to NW (if any)   + The associated procedure   + Usage/Applicable use case(s) of MI-Option 1   Note: whether MI-Option 1 is needed or not is a separate discussion  **Agreement #116bis**  From RAN1 perspective, for UE-sided model(s) developed (e.g., trained, updated) at UE side, following procedure is an example (noted as AI-Example1) of MI-Option1 for further study (including the feasibility/necessity)  o A: For data collection, NW signals the data collection related configuration(s) and it/their associated ID(s)  § Associated IDs for each sub use case in relation with NW-sided additional conditions  o B: UE(s) collects the data corresponding to the associated ID(s)  o C: AI/ML models are developed (e.g., trained, updated) at UE side based on the collected data corresponding to the associated ID(s).  o D: UE reports information of its AI/ML models corresponding to associated IDs to the NW. Model ID is determined/assigned for each AI/ML model  § relationship between model ID(s) and the associated ID(s)  § How model ID(s) is determined/assigned, e.g.,  · Alt.1: NW assigns Model ID  · Alt.2: UE assigns/reports Model ID  · Alt.3: Associated ID(s) is assumed as model ID(s)  □ “Model ID is determined/assigned for each AI/ML model” in D is not needed  · Alt.4: Model ID is determined by pre-defined rule(s) in the specification  § FFS: how to report  § Note: D is to facilitate AI/ML model inference  o Note: Step A/B/C and additional interaction of associated IDs between UE and NW can be considered as a different solution for resolving the consistency without model identification.  **Agreement RAN#117**  **From RAN1 perspective, for UE part of two-sided model, further study the following example of MI-Option2 (including the feasibility/necessity)**   * **AI-Example2-1** * **A: A dataset is transferred from the NW/NW-side to UE/UE-side via standardized signaling.**    + **Note: RAN1 study of Step A only focuses on RAN1 aspect of the dataset transfer from NW to UE. Other solution for dataset exchange is out of RAN1 scope.** * **B: UE part of two-sided model(s) is(are) developed based on at least the above dataset.** * **C: UE reports information of its UE part of two-sided model(s) corresponding to the above dataset to the NW.** * **FFS: How model ID is determined/assigned for each AI/ML model (including relationship between dataset and model ID)** * **Note: Some step(s) may not be needed for MI-Option2** * **Note: The above example is based on the assumption of NW-first training. It is separate discussion for the assumption of UE-first training.** * **Note: The study should consider the impact on inter-vendor collaboration, at least including complexity, performance, interoperability in RAN4/testing related aspects and feasibility.** * **FFS: whether/how to consider UE-side additional condition(s) for the dataset** |

In RAN#117, it is agreed to further study on how model is identified for MI-Option2 (including the feasibility/necessity). When dataset is transferred from the NW to UE in Step A of AI-Example 2-1 in the above agreement, the association between the model ID and the dataset is indeed straightforward. The base station provides the dataset to the UE, along with a model ID. The UE utilizes the dataset to train its UE-sided model. This option can be applied to both UE-side model and two-sided model since the dataset includes all the information the UE needs to train the corresponding model.

**Proposal 1: NW assigns the Model ID associated with the dataset and this is transferred from NW to UE (as shown in AI-Example2-1 step A).**

In previous meetings discussions happened on AI-Example1 of MI-Option1 about the associated ID. In our view, global ID is unique across all cells, while local ID is unique within a group consisting of one or multiple cells. The global ID provides consistency across the network while in local ID (per cell), NW needs to change the local ID when UE changes the serving cell, which will increase the NW management burden. If ID is global, NW needs to manage a single ID, so which is consistent across the entire network.

**Proposal 2: Regarding AI-Example1 of MI-Option1,**

* **Global Cell Identity (GCI) can be used as an associated ID.**

In RAN#116-bis it is discussed that How model ID(s) is determined/assigned in , three alternatives are provided model ID determination i.e.;

Alt.1: NW assigns Model ID

Alt.2: UE assigns/reports Model ID

Alt.3: Associated ID(s) is assumed as model ID(s)

□ “Model ID is determined/assigned for each AI/ML model” in D is not needed

Alt.4: Model ID is determined by pre-defined rule(s) in the specification

We prefer Alt1, where the Network (NW) assigns the Model ID. In this scenario, the dataset is transferred to the NW. The NW will then determine the appropriate Model ID based on the dataset and communicate it to the User Equipment (UE).

For AI-Example1 of MI-option1 the relationship between model ID(s) and the associated ID(s) is considered as follows. One associated ID can be linked to multiple Model IDs (for e.g., Local Model and Global model).

**Proposal 3: Regarding the relationship between model ID(s) and the associated ID(s) in AI-Example1 of MI-Option1, one associated ID(s) can be linked to multiple model IDs.**

**Proposal 4: Regarding AI-Example1 of MI-Option1**

* **Down select to Alt1 (i.e., NW assigns Model ID)**

# Model transfer/delivery

In TR 38.843 [1], the following observations for model transfer/delivery cases are provided.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4.3 Collaboration levels  …..  Table 4.3-1: Model delivery/transfer cases   |  |  |  |  | | --- | --- | --- | --- | | **Case** | **Model delivery/transfer** | **Model storage location** | **Training location** | | **y** | model delivery (if needed) over-the-top. | Outside 3GPP Network | UE-side / NW-side / neutral site | | **z1** | model transfer in proprietary format. | 3GPP Network | UE-side / neutral site | | **z2** | model transfer in proprietary format. | 3GPP Network | NW-side | | **z3** | model transfer in open format. | 3GPP Network | UE-side / neutral site | | **z4** | model transfer in open format of a *known model structure* at UE, i.e., an exact model structure as has been previously identified between NW and UE and for which the UE has explicitly indicated its support. | 3GPP Network | NW-side | | **z5** | model transfer in open format of *an unknown model structure* at UE, i.e., any other model structure not covered in z4, including any model structure that is only partially known. | 3GPP Network | NW-side | | Note: The definition of various Cases is only for the purpose of facilitating discussion and does not imply applicability, feasibility, entity mapping, architecture, signalling nor any prioritization. | | | |   When a model of a known structure at UE (e.g., Case z4) is transferred from the Network, the new model being identified (e.g., via Type B2) has the same structure as a previously identified model at the Network and UE.  For model delivery/transfer to UE (for UE-side models and UE-part of two-sided models):   * Model delivery/transfer to UE, if feasible, may be beneficial to handle scenario/configuration specific (including site-specific configuration/channel conditions) models (i.e., when a single model cannot generalize well to multiple scenarios/configurations/sites), to reduce the device storage requirement. * Model delivery/transfer to UE after offline compiling and/or testing may be friendlier from UE’s implementation point of view compared to the case without offline compiling and/or testing. On the other hand, the case without offline compiling and/or testing (that can update parameter with known model structure), may have benefit at least in terms of shorter model parameter update timescale. * Model transfer/delivery of an unknown structure at UE has more challenges related to feasibility (e.g. UE implementation feasibility) compared to delivery/transfer of a known structure at UE. * For model trained at network side, Case y (w/ NW-side training) and Case z2 may incur the burden of offline cross-vendor collaboration such as sending a model to the UE-side and/or compiling a model. * For model trained at UE side/neutral site, Case z1 and Case z3 may incur the burden of offline cross-vendor collaboration to send the trained model from the UE-side to the network, compared to Case y (w/ UE-side training) which does not have such burden. * Model storage at the 3GPP network, compared to storing the model outside the 3GPP network, may come with 3GPP network side burden on model maintenance/storage.   Proprietary design disclosure concern may arise from model training and/or model storage at the network side compared to other cases (such as case y with UE side training) which does not have such issue. |

The table 4.3-1 in TR 38.843 identified the entities that will be used as training location and model storage location for UE side and UE part models.

The following agreement is agreed in RAN#117 for Model transfer/delivery

|  |
| --- |
| **Agreement**  **From RAN1 perspective, for model delivery/transfer Case z4, further study the following alternatives (including the necessity/feasibility/benefits):**   * **Alt. A**   + **Step A-1: UE reports the supported known model structure(s) to network**   + **Step A-2: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step A-1**   + **FFS: whether some additional step(s), and/or whether other information is needed** * **Alt. B**    + **Step B-0: UE reports to NW its support of model transfer/delivery case z4**     - **Note: Step B-0 may be before or after Step B-1, or not necessary**   + **Step B-1: NW indicates to UE the candidate known model structure(s)**   + **Step B-2: UE reports to NW which model structure(s) out of the candidate known model structure(s) indicated in Step B-1 is supported**   + **Step B-3: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step B-2**   + **FFS: whether some additional step(s), and/or whether other information is needed** * **Note: Other alternative(s) is not precluded** * **Note: Other method(s) of parameter exchange from NW to UE side is a separate discussion.** |

It has been agreed that two alternatives for model delivery/transfer for Case z4. In Alt A, UE reports the supported known Model structure to the network upon receiving the Model structure NW transfers parameters to the UE. In Alt. B, UE reports to NW its support for model transfer/delivery. Upon receiving support from UE, NW transmits known models to the UE. The UE reports support for known model to the NW, upon receiving known model from the NW, NW transmits corresponding parameter to the UE. In our view Alt. B increases burden on both NW and UE side with extra signaling we prefer Alt1 over Alt2 with one more additional step as Step A-3 as based on received parameter, the UE compiles and tests if needed.

**Proposal 5: Consider Alt1 for model delivery/transfer in Case z4**

* **Alt. A**
  + **Step A-1: UE reports the supported known model structure(s) to network**
  + **Step A-2: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step A-1**
  + **Step A-3: based on received parameters, the UE compiles and tests if needed.**
  + **FFS: whether some additional step(s), and/or whether other information is needed**

**Proposal 6: For model delivery/transfer Case z4 Alt. A should be prioritised.**

In model transfer is in open format of a *known model structure* at UE, i.e., an exact model structure as has been previously identified between NW and UE and for which the UE has explicitly indicated its support. For open format model delivery/transfer Case z4 reuse the existed open format in AI like Open Neural Network *Exchange (*ONNX).

**Proposal 7: For the open format for model delivery/transfer Case z4, reuse the existing open format(s) that is already existing in the AI community (for e.g., ONNX).**

# Conclusion

In this contribution, we discussed study aspects related to model identification and model transfer/delivery. In particular, the contribution has the following proposals.

**Proposal 1: NW assigns the Model ID associated with the dataset and this is transferred from NW to UE (as shown in AI-Example2-1 step A).**

**Proposal 2: Regarding AI-Example1 of MI-Option1,**

* **Global Cell Identity (GCI) can be used as an associated ID.**

**Proposal 3: Regarding the relationship between model ID(s) and the associated ID(s) in AI- Example1 of MI-Option1, one associated ID(s) can be linked to multiple model IDs.**

**Proposal 4: Regarding AI-Example1 of MI-Option1**

* **Down select to Alt1 (i.e., NW assigns Model ID)**

**Proposal 5: Consider Alt1 for model delivery/transfer in Case z4**

* **Alt. A**
  + **Step A-1: UE reports the supported known model structure(s) to network**
  + **Step A-2: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step A-1**
  + **Step A-3: based on received parameters, the UE compiles and tests if needed.**
  + **FFS: whether some additional step(s), and/or whether other information is needed**

**Proposal 6: For model delivery/transfer Case z4 Alt. A should be prioritised.**

**Proposal 7: For the open format for model delivery/transfer Case z4, reuse the existing open format(s) that is already existing in the AI community (for e.g., ONNX).**

# References

1. RP-234039, “New WID on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface”, 3GPP TSG RAN Meeting #102, Edinburgh, Scotland, December 11-15, 2023.
2. TR38.843, “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface (Release 18)”, V18.0.0, December 2023