**3GPP TSG RAN WG1 #120 R1-250XXXX Athens, Greece, February 17th – 21st, 2025**

Source: Moderator (OPPO)

Title: Summary#2 for other aspects of AI/ML model and data

Agenda Item: 9.1.4.2

Document for: Discussion and Decision

# Introduction

Rel-19 work item on AI/ML for NR air interface was approved as RP-213599 in RAN#102. Generally, the Rel-19 AI/ML WID includes two categories of objectives:

* Normative work for basic AI/ML general work, AI-based management, AI-based positioning
* Study of some controversial topics / advanced features, e.g., AI-based CSI, model identification, training data collection for UE-sided model, model transfer/delivery

Accordingly, RAN1 chair arranged several agenda items for different topics, among which this agenda item focuses other aspects of AI/ML model and data including model identification/procedure, training data collection for UE-sided model, and model transfer/delivery. The corresponding objectives were captured in the Rel-19 WID (RP-213599) and the WID were further updated to RP-242399 in RAN#105 meeting. The counterparts are copied as below for reference:

|  |
| --- |
| Study objectives ~~with corresponding checkpoints in RAN#105 (Sept ’24)~~:   * … * Necessity and details of model Identification concept and procedure in the context of LCM for two-sided models [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 |

In this summary, the key ideals and proposals from companies are summarized, and offline proposals are drafted based on company contributions for further discussion.

Regarding the file names, companies are encouraged to follow the guidance of R1-2203012 (Page 16) as below:

|  |
| --- |
| * + - To avoid ending-up with too long file names and downloading/opening issues, the following naming convention is recommended:       * Keep the previous company’s name (only the most recent one) in the filename, e.g.         + 5/Summary-1-v000-Moderator (HW)         + 5/Summary-1-v001-LG         + 5/Summary-1-v002-LG-CATT         + 5/Summary-1-v003-CATT-vivo         + 5/Summary-1-v004-Moderator(HW)       * It helps identifying on which previous version your input is based on and solve any crossing emails issue. Note the use of 3digit version numbers in the file names. |

# Model identification/procedure

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI [1] | *Observation 1: Associated ID across multiple cells is beyond the scope of RAN1 discussion.*  *Proposal 1: For MI-Option 1, conclude that an associated ID is valid only within a cell, and the network assigns/manages associated IDs.*  *Observation 2: For MI-Option 1, one associated ID may be mapped to multiple models trained using the data collected based on the same associated ID.*  *Proposal 2: Conclude that associated ID is not model ID.*  *Proposal 3: Clarify the relationship between model ID and the corresponding dataset used for model training, in particular, the method of identifying a model based on the transferred dataset for model training.*  *Proposal 4: For MI-Option 2, conclude that UE-side additional condition(s) do not need to be considered for UE part of two-sided model.*  *Proposal 5: For MI-Option 2, conclude that model IDs are assigned only by the NW.*  *Proposal 6: For MI-Option 2, conclude that ID-X is a dataset ID, not a model ID.*  *Proposal 7: Support MI-Option 3 with further study of its procedures and specification impact, based on model transfer Case z4.*  *Proposal 8: A general procedure for MI-Option 3 includes the following steps:*   * *Step 0: The network obtains information about UE’s capability of model inference.* * *Step A: The network develops AI/ML models considering UE capability.* * *Step B: The network transfers the developed model with a model ID to the UE.* * *Step C: The UE confirms model transfer or delivery.*   *Proposal 9: RAN1 to discuss and decide whether MI-Option4 is a valid option for model identification.*  *Proposal 10: All options of MI-Option 4, i.e., Case-MI-4A, 4B and 4C, should be studied.*  *Proposal 11: Clarify the relationship between the standardized reference model and possible multiple derived models. For example, whether all derived models can share the same ID of the reference model.*  *Proposal 12: If UE/UE-side develops multiple models compatible to the same reference model, all these derived models need to be identified by network, if the network needs to differentiate them (e.g., for performance monitoring).*   * *The derived models can be identified using any approved model identification options.*   *Proposal 13: Study the feasibility of supporting two-sided models without model identification.* |
| ZTE [2] | *Observation 1: Regarding MI-Option2, the feasibility is questionable due to*   * *Huge resource overhead consumption on over-the-air dataset exchange* * *Potential performance degradation and interoperability problem in actual deployment.* * *Large latency on model deployment timescale* * *Large UE power consumption for receiving dataset*   *Proposal 1: Regarding MI-Option 2, dataset ID is considered as model ID.*  *Proposal 2: Regarding MI-Option 2, RAN1 further studies the necessity and potential approaches (if needed) to deal with the impact of UE-side additional conditions for the dataset.*  *Proposal 3: Regarding MI-Option 2, detailed mechanism of dataset transfer is up to RAN2.*  *Observation 2: Regarding MI-Option3, the feasibility can be achieved due to*   * *Less resource overhead consumption on over-the-air model transfer* * *Minor performance degradation via model transfer during actual deployment* * *Moderate latency on model deployment timescale*   *Observation 3: Regarding MI-Option 4, there is no such issue as multi-vendor collaboration and model pairing if reference UE-part model is standardized.*  *Proposal 4: Regarding MI-Option 4, standardization of reference UE-part model is preferred.*  *Proposal 5: In Rel-19 AI/ML framework study, type B model identification is prioritized compared with type A model identification.*  *Observation 4: The comparison among MI-Option2, MI-Option3 and MI-Option 4 is as following.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | *Category* | *Model identification* | *Multi-vendor collaboration* | *Whether model pairing is addressed* | *Analysis* | | *Dataset* | *MI-Option 2* | *Option 4-1* | *Yes* | *MI-Option 2 (dataset transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue. However, the problem of dataset transfer needs to be further addressed.* | | *Model transfer* | *MI-Option 3* | *Option 3a-1* | *Yes* | *MI-Option 3 (model transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue. However, the problem of model transfer needs to be further addressed.* | | *Standardization of reference models* | *MI-Option 4* | *Option 1* | *Yes* | *MI-Option 4 (standardized reference model) can be applied to address the multi-vendor collaboration issue and model pairing issue. However, the problem of reference model standardization needs to be further addressed.* | |
| Huawei [3] | *Observation 1: If MI-Option 4 needs to be classified to model identification, the definition of model identification may need to be revisited.*  *Proposal 1: There is no need to further discuss MI-Option 1 in 9.1.4.2, since MI-Option 1 is applicable only to one-sided model cases, while the scope of model identification in revised WID is limit to two-sided case, for which the UE side does not need to be provided with NW-side additional condition since the data collection is performed by NW side for NW-side training or NW-first training.*  *Proposal 2: For the transmitted information of MI-Option 2, if the dataset is delivered from NW side to UE side, the following information may be needed:*   * *Content of model input and label for training the UE side CSI generation part.* * *Other meta information, including at least: dataset ID, size of dataset, type/format of data samples, association between ground-truth CSI and CSI feedback subject to a data sample, scalability information, quantization method for CSI feedback, dataset split/segmentation information, and target performance information.*   *Proposal 3: For the procedure of MI-Option 2, the model identification, the “ID-X” could be interpreted as dataset ID.*  *Proposal 4: For the transmitted information of MI-Option 3, taking Case z4 for example, the following information may be needed:*   * *Model parameters.* * *Other meta information, including at least: model ID, format of the parameters, model structure information, quantization method and parameters.*   *Proposal 5: For the procedure of MI-Option 3, the model identification is achieved when the model ID is delivered in together with the delivered model.* |
| Spreadtrum [4] | *Observation 1: From RAN1 perspective, the following procedure is an example of MI-Option 1 for two-sided model:*   * *Step 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* * *Step B: UE(s) collects the data corresponding to the associated ID(s)* * *Step C: UE(s) reports the collected data to the NW* * *Step D: AI/ML models are developed (e.g., trained, updated) at NW side based on the collected data corresponding to the associated ID(s).* * *Step E: AI/ML models or data are delivered to UE by NW with corresponding model ID* * *Note: It is up to NW to assign the model ID.*   *Proposal 5: MI-Option 1 shall not be considered for two-sided use case.*  *Proposal 6: MI-Option 2/3/4 can be considered for two-sided use cases.* |
| CATT [5] | *Observation 1: MI-Option1 is out of scope, since AI-Example1 is only applicable for UE-sided model that developed at UE side, but the revised WID clearly states that model identification is only studied in two-sided model use case.*  *Observation 2: Even though the UE-side additional condition should impact the performance theoretically, RAN1 didn’t (or at least insufficiently) evaluate and never identify a specific UE-side additional condition that has to concern in AI/ML-based CSI compression.*  *Proposal 1: No need to discuss MI-Option 1 or associated ID in this agenda.*  *Proposal 2: Whether other information than ID-X is needed for pairing depends on:*   * *How dataset is constructed;* * *How many ID-X(s) is assumed to be transmitted to along with the dataset;* * *The effective range of an ID-X, e.g. whether ID-X can be used across different cells.*   *Proposal 3: For a transmitted dataset, NW can provide information on how the dataset is constructed to UE/UE-side, at least including from which cells the data samples are collected.*  *Proposal 4: In MI-Option2, study applicable range of ID-X to clarify whether dataset can be uniquely identified across different cells.*  *Proposal 5: In AI-Example2-1, as a starting point, the mapping relationship between ID-X and model ID is flexible, i.e. not limited to one-on-one mapping.*   * *The model ID is assigned by network after the UE reports information of its UE part of two-sided model(s);* * *FFS the prerequisite when ID-X and model ID is one-one-one mapping, and how to determine model ID in this case.*   *Proposal 6: In AI-Example2-1, for AI/ML-based CSI compression, the need and benefit of UE-side additional condition(s) is unclear.* |
| China Telecom [6] | *Proposal 5: Clarification on the relationship between model ID and ID-X is needed.* |
| CMCC [7] | *Proposal 1: The following aspects could be the starting point when discussing the information of model during model identification:*   * *The related functionality or AI enabled feature* * *Model’s applicable scenarios, configurations* * *Type/dimension of model input/output*   *Proposal 2: There may be one dataset ID associated with the transferred dataset in the Step A in MI-Option2 and the dataset ID can be interpreted as a kind of ID-X.*  *Proposal 3: It is suggested to further study the following two alternatives for model ID(s) determination/assignment in MI-Option2:*   * *Alt.1: NW assigns model ID* * *Alt.2: UE assigns/reports model ID*   *Proposal 4: For MI-Option 2, the following meta information may be needed before/during dataset transfer:*   * *Input and output of the CSI generation part and/or the CSI reconstruction part* * *Type/format of data samples* * *Model scalability information* * *Quantization method for CSI feedback* * *Backbone of model*   *Proposal 5: Some necessary model related information, such as model backbone, could be aligned between NW side and UE side to achieve better performance for MI-Option2.*  *Proposal 6: Considering the performance gain, specification efforts and fact that RAN4 is studying the issue, MI-Option 4 for model identification can be deprioritized at current stage.* |
| vivo [8] | *Observation 1: Global associated ID may expose deployment choices of NW side, but is useful information to maintain consistency between training and inference.*  *Observation 2: Local associated ID either requires huge or infeasible efforts at UE side to categorize the collected data or may require cell/site/region specific model development and management.*  *Proposal 1: Local associated ID for multiple cells can be supported.*   * *Local associated ID for multiple cells is useful for maintaining consistency between training and inference and help UE to train a model with good generalization performance, for a larger area than a single cell.* * *Local associated ID for multiple cells may expose less deployment choices of NW side, than global associated ID.*   *Proposal 2: Associated ID + cell ID(s) can be supported to indicate the applicable cell(s) for multiple cell scenario.*  *Proposal 3: Regarding the associated ID for Rel-19,*   * *The UE-side burden at least including burden on data collection, training, model delivery/management and power consumption may be reduced if the UE can assume that NW-side additional conditions with the same associated ID are consistent among multiple cells.* * *Information on mapping between NW-side additional conditions containing proprietary information to an associated ID should not be disclosed to other vendor(s).* * *It may incur burden of NW including complexity and configuration constraints if the associated ID is used to ensure the consistency among multiple cells if number of cells is large. Otherwise, there should not be such concerns for NW.* * *[NW proprietary information may be disclosed if the associated ID is used to ensure the consistency among multiple cells if number of cells is large. Otherwise, there should not be such concerns for NW.]* * *Note: Feasibility/details of the mechanism(s) is discussed per use case*   *Proposal 4: Associated ID + time stamp information can be supported to indicate the applicable period of the associated ID.*  *Proposal 5: ID-X is model ID (reusing the same ID as MI-Option3) or associated ID (reusing the same ID as for AI/ML for beam management), for the purpose of unified MI framework.*  *Observation 3: Feasibility of model identification with dataset transfer is dependent on the feasibility of dataset transfer itself.*  *Proposal 6: Model identification is needed for cases where multiple models are transferred from NW to UE.*  *Proposal 7: For MI-Option4, if multiple reference models are standardized, an ID can be pre-defined for each reference model:*   * *The model(s) can be identified by its pre-defined ID at UE/network.*   + *FFS: details of the pre-defined IDs.* * *FFS: if UE/UE-side develops multiple models compatible to the same reference model, whether these models are needed to be identified by network or not?* * *Note: whether multiple reference models can be standardized or not is a separate discussion.*   *Proposal 8: Model identification via standardization of reference models may have the following procedures:*   * *MI-Option 4-1: UE may report specified (global) model ID of reference model. Specified (global) model ID is used for model control and performance monitoring.* * *MI-Option 4-2: UE may report specified (global) model ID of reference model. Then NW assigns local model ID for specified (global) model ID. Local model ID is used for model control and performance monitoring.*   *Proposal 9: Reference model may be also used in one-sided case. For example, RAN4 may also define some reference model for one-sided case.*  *Proposal 10: MI-Option 4 (model identification via standardization of reference models) can be used in cases when multiple reference models are specified, which would have the following purpose/usage.*   * *Would partially ensure consistency between training and inference, where multiple reference models are specified considering more additional conditions from vendors;* * *Can support different AI model with different capabilities, if multiple reference models with different capabilities are pre-defined.* |
| Ericsson [9] | *Proposal 1 RAN1 to conclude that there is no need to further discuss the MI-Options in the other aspects agenda item, further progress can be part of the two-sided CSI compression use case.* |
| OPPO [10] | *Proposal 1: Support a unified LCM providing both functionality-based and ID-based operations.*   * *Functionality-based operation is supported by default, in which the granularity of the functionalities is aligned with the Feature/FG in a UE capability report, i.e., conditions.* * *Model ID can be used on top of functionality for indication of different additional conditions, to support multiple scenarios, configurations, sites, etc.*   *Proposal 2:*  *Regarding the associated ID for Rel-19,*   * *The UE-side burden at least including burden on data collection, training, model delivery/management and power consumption may be reduced if the UE can assume that NW-side additional conditions with the same associated ID are consistent among multiple cells.* * *Information on mapping between NW-side additional conditions containing proprietary information to an associated ID should not be disclosed to other vendor(s).* * *It may incur burden of NW including complexity and configuration constraints if the associated ID is used to ensure the consistency among multiple cells ~~to manage the associated IDs across cells~~.* * *The commonality between multiple cells may be implied ~~NW proprietary information may be disclosed~~ if the associated ID is used to ensure the consistency among multiple cells* * *Note: Feasibility/details of the mechanism(s) is discussed per use case*   *Proposal 3:*  *For model identification type B MI-Option 1,*   * *Step D should be supported for the UE not involved in Step A, B and C.*   + *Alt.1: NW assigns Model ID is preferred because it supports model identification for UE involved or not involved in Step A, B and C.*   + *Alt.2 is not preferred unless advantage over Alt.1 can be justified.*   + *Alt.3 is not preferred because it only supports model identification for UE involved in Step A, B and C.*   + *Details needs to be clarified for Alt.4.* * *Strive for achieving the 1-to-1 mapping between model ID(s) and the associated ID(s), thus for the same inference behavior for UE involved or not involved in Step A, B and C.*   *Proposal 4:*  *For model identification type B MI-Option 2,*   * *ID-X is the Model ID. Or, ID-X is 1-to-1 mapped to Model ID(s).* * *NW assigns Model ID in Step A.* * *Step C is needed if the UE-part of the model would be also used for UEs not involved in the model development.*   *In Step C, UE reports the information about the UE-side additional condition(s) for training the UE-part of the model to NW.*  *Proposal 5:*  *For model identification type A,*   * *Model ID is allocated to the model as well as the additional conditions used to train the model via OTT inter-vendor engineering.*   *Proposal 6: Support Global Model ID and Local Model ID*   * *Global Model ID is used for model transfer and test certification* * *Local Model ID is used for inference procedure, e.g., for indication of model selection/activation/deactivation/switching/fallback, which is similar to the resource/configuration ID in the legacy NR specification and does not include explicit information about the model, e.g., scenarios/configurations/sites.*   *Proposal 7: Regarding the relationship of model ID, first indication, and second indication for model transfer/delivery Case z4,*   * *Global Model ID consists of the information of the first indication and the second indication, e.g., Global Model ID is a combination of the first and second indications.* * *A Local Model ID can be assigned by network and is separated from the first indication and the second indication for inference procedure.* |
| Google [11] | *Proposal 1: For MI type A, it is assumed that the indication of a model ID is known by the NW and UE after UE connected to the NW.*   * *No additional specification work is required to maintain the same understanding between the NW and UE on the indication of a model ID.*   *Proposal 2: MI-Option 1 is necessary to assist the NW and UE to maintain the same understanding for the property of model input and model output, so that the NW can configure corresponding DL RS for the UE to identify the model input and configure corresponding UL resource for model output report.*  *Proposal 5: The ID-X for MI-Option2 is model ID.*  *Proposal 6: For two-side model, support the following configurations for the agreed IDs*   * *Support to configure the model ID in CSI report configuration for inference* * *Support to configure the associated ID for CSI-RS resource set for CSI acquisition for UE-side data collection*   *Proposal 7: For one-side model, support the following configurations for the agreed IDs*   * *Support to configure the associated ID in CSI report configuration for inference* * *Support to configure the associated ID for SSB/CSI-RS resource set for BM or CSI-RS resource set for CSI acquisition for UE-side data collection* * *Model ID is not applicable for one-side model* |
| TCL [12] | *Observation 1: The relationship between the other IDs and the model ID needs to be clarified. If the model ID represents model type, structure and other abstract features, i.e., it does not uniquely identify an AI/ML model, then the consistency cannot be guaranteed by model ID alone. We need other IDs together with the model ID to ensure consistency.*  *Observation 2: If the model ID represents a specific model, the NW should maintain a large number of model IDs training with different data, which will introduce additional overhead.*  *Observation 3: The ID-X can be designed according to the additional conditions group to reduce the overhead and provide clear mapping to physical parameters.*  *Proposal 1: Defer the discussion of the role of ID-X in the two-sided model until there is a consensus on the definition of ID-X.*  *Proposal 2: Support option 1 as a baseline, and option 3 is for FFS. Do not support option 2 unless further clarification is made.*  *Proposal 3: The functionality ID corresponds to the use cases or sub use cases. The model ID corresponds to a model type or model structure, not a specific model trained by a specific dataset.*  *Proposal 4: The ID-X and the dataset ID are functional overlap and can be mutually substituted, they may not be duplicated defined.*  *Proposal 5: If the ID-X is globally or cross-cell defined, an explicit mapping between the ID-X and the NW additional conditions should be defined.* |
| LG [13] | *Proposal#1. Clarify that NW can configure NW-side additional condition to UE for helping UE to select applicable functionality among supported functionalities, and the exact configuration method including whether or not to use the associated ID is up to sub-use-case signaling design.*  *Proposal#2. Details of associated ID, including whether it is needed, the name of the ID, the information inferred from the ID, how the ID is assigned/used, should be discussed and decided per sub-use-case.*  *Proposal#3. For MI-Option 2, clarify that ID-X alone cannot serve model identification purpose, and for the pairing purpose, ID-X is sufficient.* |
| NEC [14] | *Observation 1: To ensure consistency, the option based on associated ID is supported. And the associated ID is determined during model identification.*  *Observation 2: MI-Option 1/2 is suitable for adoption if Option-4 (Standardized data / dataset format + Dataset exchange between NW-side and UE-side) is agreed to address inter-vendor collaboration for CSI compression use case.*  *Observation 3: MI-Option 3 is suitable for adoption if Option-3b/5b (Parameters/models received at the UE are directly used for inference at the UE without offline engineering, potentially with on-device operations) is agreed to address inter-vendor collaboration for CSI compression use case.*  *Observation 4: A combination of MI-Option 1/2 and MI-Option 3 is suitable for adoption if Option-3a/5a (Parameters/models received at the UE or UE-side goes through offline engineering at the UE-side) is agreed to address inter-vendor collaboration for CSI compression use case.*  *Observation 5: Ensuring consistency of additional conditions using monitoring procedure results in high delay in identification of the suitable AI/ML model to run at UE, during which system performance suffers.*  *Observation 6: Not every difference in NW-side additional conditions requires a data collection configuration with a separated associated ID.*  *Observation 7: If the associated ID is only for UE assumption of the consistency within a cell, it may introduce unnecessary complexities for UE to train and to maintain a large number of models. If the associated ID is valid for UE assumption of the consistency within PLMN, it may introduce unnecessary difficulties considering inter-vendor coordination.*  *Proposal 1: Clarify the support of model identification for one-sided model use cases, at least for determining the associated ID to ensure the consistency between model training and model inference.*  *Proposal 2: RAN1 to select the model identification procedure based on the outcome of the study to address inter-vendor collaboration issues for CSI compression use case.*  *− MI-Option 1/2 to be adopted if Option-4 (Standardized data / dataset format + Dataset exchange between NW-side and UE-side) is agreed to address inter-vendor collaboration issues for CSI compression use case.*  *− MI-Option 3 to be adopted if Option-3b/5b (Parameters/models received at the UE are directly used for inference at the UE without offline engineering, potentially with on-device operations) is agreed to address inter-vendor collaboration issues for CSI compression use case.*  *− A combination of MI-Option 1/2 and MI-Option 3 to be adopted if Option-3a/5a (Parameters/models received at the UE or UE-side goes through offline engineering at the UE-side) is agreed to address inter-vendor collaboration issues for CSI compression use case.*  *Proposal 3: One or more associated ID(s) can be attached to one same model ID to reflect different NW side additional conditions.*  *Proposal 4: For inference for UE-side models, to ensure consistency between training and inference regarding NW-side additional conditions (if identified), the following options should be considered as priority:*  *− Model identification to achieve alignment on the NW-side additional condition between NW-side and UE-side*  *− Model training at NW and transfer to UE, where the model has been trained under the additional condition*  *− Information and/or indication on NW-side additional conditions is provided to UE*  *Proposal 5: To ensure the consistency within a cell and across multiple cells, support UE to feedback whether associated ID is needed, at least for model inference.*  *Proposal 6: Study the grouping of cells that can ensure the consistency within a subset of cells.* |
| Lenovo [15] | *Observation 1: The data collection related configuration(s) and/or indication(s) provided in MI-Option 1 and/or dataset transferred in MI-Option 2 is not enough to ideally identify an AI/ML model without other associated information.*  *Observation 2: There is common interacted information between the model identification procedure and inter-vendor training collaboration on the requirements on the data/dataset and model-related information sharing and indications.*  *Proposal 1: The information of UE part of two-sided model(s) reported from UE in Step D of AI-Example1 (MI-Option1) and Step C of AI-Example2-1 (MI-Option2) need to include the indication of the model structure.*  *Proposal 2: Study the necessity and feasibility of the AI/ML model-related information and/or configurations shared in the model identification procedure to support the inter-vendor training collaboration in this agenda.*  *Proposal 3: When using ID-X for the dataset to assist exchanging the information for model training/updating, the relation between ID-X and data collection related configuration(s) for dataset construction need further study.* |
| NVIDIA [16] | *Proposal 1: Conclude that there is a need for model identification in the context of LCM for two-sided models.*  *Proposal 2: Besides MI-Option 1 and MI-Option 2, describe examples for the following options to study their feasibility/necessity:*   * *MI-Option 3: Model identification in model transfer from NW to UE* * *MI-Option 4: Model identification via standardization of reference models. (for CSI compression)* * *MI-Option 5: Model identification via model monitoring* |
| Xiaomi [17] | *Observation 6: Compared with approach of step A/B/C and additional interaction of associated IDs between UE and NW, MI-Option 1 is still beneficial considering the following aspects*   * *Potential processing interruption management* * *Reducing network burden in handling the additional condition*   *Observation 7: MI-Option 1 is applicable to one-sided model*  *Observation 8: It is more efficient to deliver the data set, align the model information and determine the model ID without over-the-air signalling*  *Observation 9: Regarding to MI-Option 2, following two options are feasible:*   * *Option 1: One UE part of two-side model is trained with one dataset, and one dataset is only used for training one UE part of two-side model;* * *Option 2: One UE part of two-side model is trained with multiple datasets, and one dataset is only used for training one UE part of two-side model.*   *Proposal 7: The associated ID is not equivalents to the model ID.*  *Proposal 8: Support cell-group unique associated ID to balance the complexity on UE side and proprietary deployment preservation on NW side.*  *Proposal 9: How to utilize the datasets for UE-part of two-side models development depends on UE implementation.*  *Proposal 10: Regarding to MI-Option 2, support cell-group unique ID-X to balance the complexity on UE side and proprietary deployment preservation on NW side.*  *Proposal 11: Regarding to MI-Option 2, an additional indication is need for paring the UE-part and the NW-part of a two-sided model, to indicate the dataset is used for initial model training, model retraining, or model fine-tune.*  *Proposal 12: Consider the following procedure for MI-Option 3*   * *Step 1: model identification from NW to UE, meta information and model ID would be shared* * *Step 2: UE confirms the model transfer or delivery* * *Step 3: Model transfer/delivery from NW to UE* * *Step 4: UE reports the model ID to indicate the availability of the model*   *Proposal 13: Regarding to MI-Option 4, support to discuss it in AI 9.1.4.1 (CSI compression).*  *Proposal 14: Consider the following procedure for Type A model identification*   * *Step 1: Data set construction* * *Option 1: Dataset is obtained via offline coordination* * *Option 2: Via data collection from UE* * *Step 2:* * *Train/Update the AI model offline* * *Step 3:* * *UE side reports the model information offline. The reported information may include model input, output, associated network additional condition, performance and potential processing time for model activation or switch* * *NW side assigns the model ID for this model to UE side offline* * *Step 4:* * *UE reports the model ID to network to indicate the availability of the model*   *Proposal 15: Associated ID can be considered for data collection for type A model identification.* |
| Apple [18] | *Proposal 1: Remaining details of MI-option 2 is discussed in 9.1.4.1.* |
| Samsung [20] | *Proposal#1: To study the necessity of MI-Option1, RAN1 to consider its application on model-level management of AI/ML operations at the UE including*   * *Timeline management for LCM operations, e.g., model inference, activation, switching* * *Network’s awareness on UE’s AI/ML processing unit and its occupancy*   *Proposal#2: RAN1 to conclude ensuring consistency on network-side additional condition between model training and inference does not necessitate model identification. Indication on network-side additional condition based on Step A/B/C of AI-Example1 and additional interaction of associated IDs between UE and NW is sufficient.*  *Observation#1: For AI-Example2-1 of MI-Option2, the network may generate and transfer dataset(s) associated to NW-side additional condition(s)*   * *For two-sided models development, NW-part of two-sided model associated with the dataset can be considered as NW-side additional condition.*   *Observation#2: For AI-Example2-1 of MI-Option2, the network may generate multiple datasets each specific to different encoder backbone assumptions but associated with the same decoder. What the network need to indicate during inference is the compatibility/pairing information, e.g., associated ID, which implicitly indicates the NW-side additional condition(s) including at least NW-part of two-sided model, rather than indication for the dataset(s).*  *Proposal#3: For AI-Example2-1 of MI-Option2, support the indication associated ID for implicit indication of NW-side additional condition(s) when dataset is transferred from the network-side to the UE-side.*   * *For NW-first two-sided models training, conclude that NW-part of two-sided model associated with the dataset can be considered as NW-side additional condition, i.e., compatibility information.* * *ID-X as agreed in RAN1#119 is the associated ID.*   *Proposal#4: For MI-Option 1 and MI-Option 2, consider the following additional procedure for model-ID-based LCM with model identification Type B1*   * *For NW’s indication on NW-side additional condition: The network provides the list of indicator(s) of network-side additional conditions for an AI/ML-enabled feature/FG* * *For UE’s model identification to the network: The UE identifies a model with information on the supported configurations/conditions for AI/ML-enabled feature/FG and/or associated indicators for NW-side additional conditions.* * *For model-ID based LCM: Network use model ID(s) for the identified model(s) to give LCM assistance, e.g., model activation, inference, monitoring, deactivation.*   *Proposal#5: For functionality-based LCM, to maintain the UE complexity in the inference phase, UE may report the maximum number of simultaneously active functionalities it supports.*  *Proposal#6: For model-ID based LCM, UE may report the maximum number of simultaneously active models it supports, the AI processing unit associated with each model, the mapping between the reported/identified models and functionalities/configurations as well as the maximum AI processing unit (APU).*  *Proposal#7: For MI-Option 4: model identification via standardization of reference models consider the following options:*   * *MI-Option 4 Type A: Model-ID identifies a fully standardized reference model or model structure* * *MI Option 4 Type B1: Model-ID indicates UE’s identified model compatible with one or more standardized reference model*   *Proposal#8: For MI-Option 4, model identification via standardization of reference model(s) or model structure(s), UE may indicate the supported AI/ML model IDs for a given AI/ML-enabled Feature/FG in a UE capability report.* |
| ETRI [21] | *Observation 1: The Associated ID can be used to configure and categorize datasets generated through the data collection process.*  *Observation 2: MI-Option1 is not applicable for the model identification for two-sided model use cases.*  *Proposal 1: For MI-Option2 for the two-sided model, ID-X and Model ID can have a many-to-many relationship.* |
| Fujitsu [22] | *Proposal-1: Regarding the ID-X of MI-Option2, the functionalities of ID-X include the following two aspects:*   * *During dataset transfer from NW to UE, it is used to indicate which NW part model is for the dataset generation* * *During model identification/report from UE to NW, it is used to indicate which NW part model the UE can jointly work with in its model inference phase*   *Proposal-2: For MI Option-2, when multiple datasets generated from the same NW part model are transferred from NW to UE, the indication of the multiple datasets should take the following issues into account:*   * *The association between the multiple datasets and the same source NW part model* * *The identification of different datasets from the same source NW part model*   *Proposal-3: Regarding MI-Option2, the following information should be transmitted from NW to UE for dataset transfer and development of the UE part model besides data samples in the dataset:*   * *Format of the data sample (e.g. input part, output part)* * *Model testing related information, e.g. test dataset information, target performance etc.* * *Dataset segment information, dataset size information* * *Dataset time-stamp information* * *Other information if needed*   *Proposal-4: A unified model identification procedure after model update based on either dataset transfer or model transfer is suggested to be studied for both MI-Option 2 and MI-Option 3:*   * *FFS: when/how to perform model identification for the new model generated by dataset/model transfer* |
| Nokia [23] | *Observation 1: Based on our analysis of MI-Option 2, the steps and the ID-X can be used for the inter-vendor collaboration Option 4-1.*  *Proposal 1: For the 2-sided models use case, RAN1 to prioritize the study on the details of the model identification MI-Option 2 steps for the purpose of the inter-vendor collaboration Direction A Option 4-1.*  *Observation 2: Based on our analysis, assuming that only the UE-part of the model structure is to be standardized in the inter-vendor collaboration Direction A Option 3 , the MI-Option 3 is applicable for model identification.*  *Observation 3: In the model transfer/delivery Case z4, Opt1 and Opt2 have lower complexity compared to Opt3. There are no additional scalability issues for Opt1 and Opt2, whereas scalability of Opt3 need further analysis.*  *Proposal 2: In the model transfer/delivery Case z4, Model ID should be derived from the first indication and the second indication when UE supports multiple model structures, and only from the second indication when the UE supports a single model structure.*  *Observation 4: Based on our analysis, assuming that only the UE-part of the model is to be standardized in the inter-vendor collaboration Option 1 (Direction C), the MI-Option 4 with the outlined Step A-D is applicable for model identification.* |
| Continental Automotive [24] | *Proposal 1: Study a more comprehensive identifier parameter for ID-X in pairing of two-sided model.*  *Proposal 2: Study a hierarchical model identification structure.*  *Proposal 3: Study a standardized model ID mapping configuration within model identification frameworks.*  *Proposal 4: Support multi-vendor AI/ML model compatibility through standardized model identification frameworks.* |
| AT&T [25] | *Proposal 1: For Rel-19, support a unified LCM providing both functionality-based and model-ID-based operations.*   * *Functionality-based operation is supported by default.* * *Model-ID, if needed, can be used in the unified LCM for model ID based LCM operations.*   *Proposal 2: Confirm the following definitions for supported functionalities, applicable functionalities and activated functionalities*   * *Supported functionalities refer to functionalities that UE can indicate by using UE capability information (via RRC/LPP signalling)* * *Applicable functionalities refers to functionalities that the UE is ready to apply for inference* * *Activated functionalities refers to functionalities already enabled for performing inference*   *Proposal 3: For all model identification options (MI-option 1, MI-option 2, MI-option 3, MI-option 4) for model identification type B,*   * *Network assigns the model ID(s) for the identified model(s) if model ID(s) assignment is needed* * *FFS: How to define a model ID for assignment*   *Proposal 4: Regarding the relationship between model ID(s) and the associated ID(s) in AI-Example1 of MI-Option1, further study the following options (including the necessity/benefit)*   * *ID-Rel-Option1: One model ID is linked to one associated ID by one-to-one mapping* * *ID-Rel-Option2: One model ID can be linked to multiple associated IDs and each associated ID is only be linked to one model ID* * *ID-Rel-Option3: One associated ID(s) can be linked to multiple model IDs and each model ID is only linked to one associated ID* * *ID-Rel-Option4: Model ID(s) can be linked to associated ID(s) by many-to-many mapping*   *Proposal 5: For AI-Example1 of MI-Option1, study and down-select the following alternatives on determining/assigning model ID (if supported).*   * *Opt.1: NW assigns Model ID* * *Opt.2: UE assigns/reports Model ID* * *Opt.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* * *Opt.4: Model ID is determined by pre-defined rule(s) in the specification*   *Proposal 6: Prioritize study of Opt.1 and Opt.3 for the assessment of Model ID, in Rel-19.*  *Proposal 7: Regarding MI-Option2 (i.e., model identification with dataset transfer) for two-sided model, where ID-X is transmitted along with the dataset, the UE-part of a two-sided model cannot be identified by the associated ID-X(s) of the dataset(s) based on which the UE-part of a two-sided model is developed.*   * *Model ID or pairing ID is required to identify the UE-part of a two-sided model.*   *Proposal 8: Regarding MI-Option2 (i.e., model identification with dataset transfer) for two-sided model, at least the following information is transmitted along with the dataset from network to UE*   * *Input data corresponding to input of UE part of the two-sided model and the labels corresponding to output of UE part of the two-sided model and their associations* * *Format/type of input data and labels (e.g., quantization information, contents, …)* * *Size of dataset (i.e., number of the pairs of (input data, label))*   + *FFS: other information* * *Note: “input data” and “label” are used for discussion purpose*   *Proposal 9: Regarding the study of Case-MI-4-1 of MI-Option4, if multiple reference models are standardized, an associated model ID can be pre-defined for each reference model*   * *The model(s) can be identified by its pre-defined associated model ID at UE/network*   + *FFS: details of the associated model IDs* * *Note: whether multiple reference models can be standardized or not is a separate discussion* |
| Qualcomm [26] | *Proposal 1: Conclude that the progress made so far including identifying different model identification options, and further details and procedures associated with those options are sufficient as far as the scope of this agenda item is concerned.* |
| NTT DOCOMO [27] | *Observation 1: For the support of scenario/site specific models, the following aspects should be considered.*  *・(Training phase) How to prepare scenario/site specific models. In other words, how to prepare models specific to additional condition.*  *・(Inference phase) How to select an appropriate scenario/site specific model among prepared models. In other words, how to ensure consistency between NW side additional conditions and UE side model.*  *Observation 2: Model identification changes management granularity from associated ID to model ID, which increases NW management burden and NW awareness of UE side performance.*  *Observation 3: MI-Option1 is useful for dataset categorization based on NW side additional condition, which helps UE side offline engineering in direction A and C.*  *Observation 4: ID-X can be the same as model ID and associated ID, where model ID and associated ID represent one NW logical model and NW side additional condition associated with NW logical model, respectively.*  *Observation 5: MI-Option3 is applicable with two-sided model and one-sided model, where the procedure of MI-Option3 can be described as follows:*  *Step1: NW side obtains the information about supportable model at UE device.*  *Step2: AI/ML models are developed and stored at NW side.*  *Step3: NW transfers the developed model with model ID.*  *Proposal 1: 3GPP should consider the framework to support scenario/site specific model.*  *Proposal 2: Future compatibility with model transfer and model storage at NW side should be taken into consideration after they are supported in 3GPP.*  *Proposal 3: Since MI-Option1 does not help two-side model pairing, MI-Option1 should be used together with other MI-Options (e.g., MI-Option 2/3/4) for two-sided model.* |

#### **Background**

During the R18 study, two types of LCM (i.e., functionality-based LCM and model-ID-based LCM) were identified. The functionality-based LCM is widely acknowledged as the basic LCM. The remaining issue is whether to support model-ID-based LCM or not, and if so, what the solution(s) is (are).

For the model-ID-based LCM, different model identification types (i.e., Type A, Type B1, Type B2) were identified for study and the corresponding outputs of R18 SI are captured in Section 4.2.2 of TR 38.843.

|  |
| --- |
| For UE-side models and UE-part of two-sided models:  - For AI/ML functionality identification  - Legacy 3GPP framework of feature is taken as a starting point.  - UE indicates supported functionalities/functionality for a given sub-use-case.  - UE capability reporting is taken as starting point.  - For AI/ML model identification  - Models are identified by model ID at the Network. UE indicates supported AI/ML models.  … 4.2.2 Model identification For *AI/ML model identification* of UE-side or UE-part of two-sided models, model identification is categorized in the following types:  - Type A: Model is identified to NW (if applicable) and UE (if applicable) without over-the-air signalling  - The model may be assigned with a model ID during the model identification, which may be referred/used in over-the-air signalling after model identification.  - Type B: Model is identified via over-the-air signalling,  - Type B1:  - Model identification initiated by the UE, and NW assists the remaining steps (if any) of the model identification  - the model may be assigned with a model ID during the model identification  - Type B2:  - Model identification initiated by the NW, and UE responds (if applicable) for the remaining steps (if any) of the model identification  - the model may be assigned with a model ID during the model identification  - Note: This study does not imply that model identification is necessary.  One example use case for Type B1 and B2 is model identification in model transfer from NW to UE. Another example is model identification with data collection related configuration(s) and/or indication(s) and/or dataset transfer. Note: Other example use cases are not precluded. Note: Offline model identification may be applicable for some of the example use cases.  Once models are identified, at least for Type A, UE can indicate supported AI/ML model IDs for a given AI/ML-enabled Feature/FG in a UE capability report as starting point. Note: model identification using capability report is not precluded for type B1 and type B2.  Model ID may or may not be globally unique, and different types of model IDs may be created for a single model for various LCM purposes. Note: Details can be studied in the WI phase. |

In RAN#105, the objective was updated to focus the study on two-sided model [RP-242399]:

|  |
| --- |
| * Necessity and details of model Identification concept and procedure in the context of LCM for two-sided models [RAN2/RAN1] |

## 1st round discussion

Similar to the previous meeting(s), it seems a common understanding among majority companies that MI-Option1 is only applicable to one-sided model(s), although one or two companies think MI-Option1 is applicable to two-sided model(s) as well. Thus, we will focus on other options for two-sided model(s) in the following discussion.

#### **Proposal 2.1**

**Proposal 2.1:**

**Agreement**

**Regarding MI-Option2 (i.e., model identification with dataset transfer) for two-sided model, the following information can be transmitted for the dataset from network to UE /UE-side**

* **Nominal input data corresponding to input of UE part of the two-sided model, the label corresponding to output of UE part of the two-sided model and their associations**
* **Format/type of nominal input data and labels ~~(e.g., quantization information, …)~~**
* **Size of dataset (i.e., number of the pairs of (nominal input data, label))**
* **Validation/testing related info~~, e.g., testing dataset information, target performance information~~**
* **FFS: other information**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | In CSI compression, Option 4-1 already having this clarification.  Anyways, it is Ok to capture this in high-level.  **Regarding MI-Option2 (i.e., model identification with dataset transfer) for two-sided model, the following information can be transmitted for the dataset from network to UE -side**   * **~~Nominal input~~ data corresponding to input of UE part of the two-sided model, the labels corresponding to output of UE part of the two-sided model ~~and their associations~~** * **~~Format/type of nominal input data and labels (e.g., quantization information, …)~~** * **~~Size of dataset (i.e., number of the pairs of (nominal input data, label))~~** * **Validation/testing related info, e.g., testing dataset information, target performance information** * **FFS: other information** |
| Panasonic | Probably use case specific discussion can be sufficient. |
| Fujitsu | Fine with the direction.   * Relates to the size of dataset information, we can add dataset segment information as well. * Dataset time-stamp information can be added as well. |
| Huawei, HiSilicon | 1) Main text is updated to UE or UE side to be consistent with RAN1 LS in #119.  2) Agree with Fujitsu that dataset segment information is added.  3) Quantization information to the label (i.e. CSI feedback) is needed.  **Regarding MI-Option2 (i.e., model identification with dataset transfer) for two-sided model, the following information can be transmitted for the dataset from network to UE or UE-side**   * **Target CSI ~~Nominal input data~~ corresponding to input of UE part of the two-sided model, the labels corresponding to output of UE part of the two-sided model and their associations** * **Format/type of Target CSI ~~nominal input data~~ and labels (e.g., quantization information, …)** * **Size of dataset (i.e., number of the pairs of (Target CSI ~~nominal input data~~, label))** * **Quantization information for the label** * **Dataset segment information** * **Validation/testing related info, e.g., testing dataset information, target performance information** * **FFS: other information** |
| TCL | Agree |
| Xiaomi | 1. For the main body, support Huawei’s update 2. Support to add the dataset segment information |
| QC | We prefer to discuss specifics regarding two-sided CSI compression in the CSI compression agenda, while this agenda discusses more general framework.  The first bullet is specific to the two-sided CSI compression use case, so it shouldn’t be included. |
| ETRI | We agree on the proposal. |
| ZTE | Although it is good to have some high level agreements, it is also vital to avoid duplicated discussion between framework and CSI compression. For example, both agendas may be discussing validation/testing related info. Since only one use case is considered for two sided model in Rel-19, we are ok to directly discuss this issue in the CSI compression agenda. |
| Samsung |  |
| CMCC | In 9.1.4.1 agenda, we are discussing CSI compression specific enhancements, it is better to have some high level two-sided model related discussion here. |
| CATT, CICTCI | Agree with the proposal.  Not sure it is proper to change it to a ‘CSI compression specific’ proposal. In our understanding this proposal serves for a general, high-level study. |
| QC2 | Can we first clarify whether the discussion is intended for two-sided CSI compression or future two-sided models? At least for two-sided CSI compression,   1. Whether dataset transfer (inter-vendor collaboration Option 4-1) is considered as a model identification should first be discussed and clarified. 2. Even if dataset transfer is considered as a model identification, the dataset transfer is between NW-side and UE-side, so it belongs to offline model identification (Type A), not online model identification (Type B). Therefore, MI-Option2 is not applicable. |
| Ericsson | Prefer to discuss this in the CSI compression use case. |
| LG | Agree with Panasonic, QC, ZTE, and Ericsson that it is better to discuss this in use-case-specific manner. |
| Spreadtrum | We think this agenda should be focus on general framework for all two-side model. If the design is only for the CSI compression case considerations, we should put the discussion in section 9.1.4.2. |
| Futurewei | If we want to discuss this in 9.1.4.2, we need to make the discussion at the high level, for example, removing some of the use case specific description, such as “quantization information”.  If we want to discuss this at details, we should discuss it at specific use cases, such as 9.1.4.1.  Right now it is a combination so it is not appropriate to be discussed at either AI. |
| Mod | A clarification for many proposals:   * AI 9.1.4.2 is for the general cases, rather than the specific design dedicated to CSI compression. Thus, CSI-compression-specific design should be avoided heer. * It is beneficial to provide a toolbox for various AI cases (including the current AI cases studied in Rel-19 and potential new cases in the future). Thus, we should be open to list the candidates that may not be needed in the current CSI compression discussion. * As for a specific use case, it can choose some tool(s) from the toolbox if needed. It can also introduce some new mechanism(s) if the current tools in the toolbox are not sufficient * From moderator’s perspective, we should avoid this kind of collision in this agenda item: One mechanism is needed in CSI compression, but we propose to not support this mechanism. Meanwhile, we should be open to discuss something beyond CSI compression (i.e., for general two-sided model)   The proposal is updated based in the inputs   * This proposal is for general two-sided model, rather than only for CSI compression. Thus, any CSI-compression-specific design are not included, e.g., quantization, target CSI, … * The name “Nominal input data” and the associations are kept as this name is convenient for the description of other sub-bullet and the association of input and label seems needed * “data segmentation information” is not included since it seem belonging to higher-layer signaling design details. It is better to discuss this in WI phase (if any) |
|  |  |
|  |  |

#### **Proposal 2.2**

In the tdocs, some companies suggest to study some other aspects, e.g.,

* Model backbone
* Whether a common NW-part is used for different datasets
* Whether two-sided model can be regarded as one kind of network-side condition
* …

**Proposal 2.2:**

**Agreement**

**Regarding MI-Option2 (i.e., model identification with dataset transfer) for the two-sided model, further study the following issue**

* **Whether/what/how some information (e.g., model backbone) of the UE-part ~~NW-part~~ of two-sided model assumed for the NW-side training should be transmitted along with the dataset**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Do not see the need of a separate proposal on this.  This backbone information or other information is already covered by the FFS in proposal 2.1. |
| Panasonic | If certain model backbone or other is specifically assumed for the specific use case of certain release, such information is not required to be transmitted. If these are flexible for the specific use case, such information is required. Therefore, to leave this for use case specific discussion would be more reasonable. |
| NTT DOCOMO | Support. |
| Fujitsu | Support. |
| Huawei, HiSilicon | Do not see the need of delivering the backbone information. UE side can autonomously select appropriate model backbone/structure based on the guidance of performance target information delivered by NW side.  Regarding “Whether a common NW-part is used for different datasets”, and “Whether two-sided model can be regarded as one kind of network-side condition”, it is NW side implementation and is transparent to UE side. |
| TCL | Agree. |
| Xiaomi | In our view, backbone information is helpful. We Support this direction |
| QC | Share similar view as Panasonic. |
| ETRI | We agree on the direction. |
| ZTE | Similar comments as for proposal 2.1, we are ok to leave this discussion to the use case level. |
| Samsung | Support this. To reply to Panasonic and Qualcomm, for Option 4-1 there has not been any discussion whether a fixed or multiple backbone will be considered. Moreover, in our contribution we showed that what is required to be shared is the assumed UE-part of two-sided model in the NW-side training and associated with dataset the NW shares.  **Regarding MI-Option2 (i.e., model identification with dataset transfer) for the two-sided model, further study the following issue**   * **Whether/what/how some information (e.g., model backbone) of the UE-part ~~NW-part~~ of two-sided model assumed for the NW-side training should be transmitted along with the dataset**   Another way is to simply capture a conclusion that the dataset transfer for a certain configuration of a feature could be backbone specific. |
| CMCC | Support this direction. |
| CATT, CICTCI | Agree. |
| Ericsson | Agree with Panasonic. |
| LG | Agree with Panasonic, QC, ZTE, and Ericsson that it is better to discuss this in use-case-specific manner. |
| Spreadtrum | Support. And we think there is no additional NW’s information needs to be shared to UE. |
| Futurewei | Agree with other companies that if we want to discuss at this level, it is better to discuss it at specific use case. |
| Mod | The proposal is updated based on Samsung’s suggestion |
|  |  |

#### **Proposal 2.3**

**Proposal 2.3:**

**Agreement**

**Regarding MI-Option4 for two-sided model, ~~down-select one or more among~~ consider the following cases for further study:**

* **Case-MI-4A: the standardized reference model is UE part of two-side model**
* **Case-MI-4B: the standardized reference model is NW part of two-side model**
* **Case-MI-4C: the standardized reference models includes both NW part and UE part of two-side models**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | For two-sided models, it is unclear how to do the down selection without assuming a use-case (CSI compression). Therefore, it is hard to make any generic agreement on this. |
| Panasonic | The conclusion can be different depending on the use case. I propose to discuss this in the use case specific. |
| NTT DOCOMO | Support. |
| Fujitsu | OK. |
| TCL | Agree |
| Xiaomi | But have similar consideration with Panasonic that, which case is selected may depend on the specific use case. From the general perspective, we’d better keep all feasible possibilities. We suggest to update the main body as follows  **Regarding MI-Option4 for two-sided model, ~~down-select one or more among~~ consider the following cases for further study from the perspective of feasibility, pros/cons:** |
| QC | Share similar view as Nokia and Panasonic. |
| ETRI | We are OK to study on this. |
| ZTE | Agree with Nokia and Panasonic. |
| Samsung | Ok. |
| CMCC | Support. |
| CATT, CICTCI | Rather than saying ‘down-select’, we suggest replace it by ‘further study the following sub options of MI-Option4’. |
| Ericsson | Agree with Nokia, Panasonic, QC, ZTE |
| LG | Agree with Nokia, Panasonic, QC, ZTE, Ericsson |
| Spreadtrum | Agree |
| Futurewei | Similar comment as in proposal 2.1; this level of discussion should be in specific use cases. |
| Mod | The proposal is updated based on the suggestions from Xiaomi and CATT |

#### **Proposal 2.4**

**Proposal 2.4:**

**Agreement**

**Regarding the study of MI-Option4 for two-sided model, if multiple reference models are standardized, an ID can be pre-defined for each reference model**

* **The model(s) can be referred to by the corresponding pre-defined ID at UE/network**
* **FFS: details of the pre-defined ID**
* **~~FFS: if UE/UE-side develops multiple models compatible to the same reference model, whether these models are needed to be identified by network or not?~~**
* **~~Note: whether multiple reference models can be standardized or not is a separate discussion~~**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Ok with the direction but need to restrict this to two-sided model. Also, we did not fully get the details covered by the last two bullets.  For example, if the reference model is referred by an ID, any further updates that NW or UE does on top of that does not matter much.  **Regarding the study of MI-Option4 for two-sided models, if multiple reference models are standardized, an ID can be pre-defined for each reference model**   * **The model(s) can be referred to by the corresponding pre-defined ID at UE/network** * **FFS: details of the pre-defined ID** * **~~FFS: if UE/UE-side develops multiple models compatible to the same reference model, whether these models are needed to be identified by network or not?~~** * **~~Note: whether multiple reference models can be standardized or not is a separate discussion~~** |
| Panasonic | If the reference models are specified jointly with the certain test condition, the ID for the reference model may be not required but the test ID might be sufficient. As this is related to how reference model is defined for the certain use case, I think use case specific discussion can be more suitable. |
| Fujitsu | If the reference model(s) is defined by RAN1, its ID is highly relied on the design of the reference model. If the reference model refers to the model(s) defined in RAN4, its ID should refer to RAN4’s design. We think whether the reference model(s) is designed/specified by RAN1 or RAN4 should be clarified first. |
| Huawei, HiSilicon | We do not see the need of assuming “multiple reference models”, since even specifying a single model could be time consuming. NW is not likely to simultaneously run multiple models with UE specific manner. |
| TCL | Agree |
| Xiaomi | This proposal is not urgent, we could discuss it when defining multiple reference models happens |
| QC | Seems like this is targeted for two-sided models, and whether or not multiple reference models are standardized, and the corresponding solution if this happens, should be discussed within the CSI compression agenda item. |
| ETRI | We agree. |
| ZTE | Agree with the intention of this proposal. Maybe we can directly keep the main bullet and the first subbullet, i.e., removing all the FFS and notes. |
| Samsung | Ok |
| CMCC | Support. As for the second FFS, if the reference model is UE part of two-sided model, then UE/UE-side developing one or multiple models is up to UE implementation, no need to be identified by NW. |
| CATT, CICTCI | OK. We think the proposal is straightforward, so as the first two sub-bullets. |
| Ericsson | Agree with the main text with addition that this is for the two-sided. We don’t need the sub-bullets. |
| LG | Agree with Xiaomi. Also, we think that this proposal has quite high dependency on the progress of CSI compression. It is better to avoid discussing such issue in this agenda. |
| Spreadtrum | Agree the proposal with Nokia’s modification. Details discussion can be next step. |
| Futurewei | Agree in general. |
| Mod | The proposal is updated according to the suggestions from several companies. |

#### **Proposal 2.5 (Placeholder)**

The relationship between options of the model identification and options of CSI compression are summarized in the following table.

|  |  |
| --- | --- |
| Model identification | Multi-vendor collaboration |
| ~~MI-Option 1~~ |  |
| MI-Option 2 | Option 2 (Deprioritized)  Option 4 |
| MI-Option 3 | Option 3  Option 5 |
| MI-Option 4 | Option 1 |
| ~~MI-Option 5~~ |  |

**Proposal 2.5 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | OK. We suggest to focus only on Option 4-1, Option 3-1, and Option 1. |
| NTT DOCOMO | In our view, MI-Option1 can be applicable to Option 3 and Option 4. For example, the data collection at UE side is necessary for training/testing at UE side in Option3a-1. Data collection can be via dataset delivery and UE measurements. Among these two approaches, UE measurement is more practical, as the privacy concern is a huge issue for the dataset delivery. In two-sided model, the dataset distribution mismatch between UE side and NW side should be avoided. Since MI-Option1 is one of the solution toward this issue with UE measurement-based data collection, we suggest MI-Option1 should include Option3. For the same reason, MI-Option1 should include Option4. |
| Fujitsu | OK. |
| Samsung | Same view as NTT Docomo. UE and network need to collect Target CSI for two-sided models. MI-Option 1 can be used in the data collection related configurations. |
| QC | All the MI-Options are for online model identification (Type A) and therefore NOT relevant for two-sided CSI compression. |
|  |  |

# Training data collection for UE-sided model

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| ZTE [2] | *Proposal 12: Regarding CN/OAM/OTT collection of UE-sided model training data, RAN1’s work can be triggered by RAN2 LS if needed, e.g., detailed data content and requirements, which can be discussed per use case.* |
| Huawei [3] | *Proposal 8: For the continued study of data collection for UE-side model training, lower the priority of the discussion at RAN1 due to the following reasons:*   * *The content for use cases have already been provided in the Rel-18 LS reply from RAN1.* * *Discussion of UE data collection mechanisms is out of RAN1 scope.* |
| Spreadtrum [4] | *Proposal 1: For data collection for UE-side model training, no discussion is needed in RAN1.* |
| CATT [5] | *Observation 3: In Rel-19, UE-side collects training data based on Solution 1a, i.e. transparent to 3GPP.* |
| CMCC [7] | *Proposal 12: Regarding the UE side data collection mechanism, RAN2 could take the Reply LS on Data Collection Requirements and Assumptions (R1-2310681) as the baseline.* |
| NVIDIA [16] | *Proposal 3: Conclude that there is a need for collection of UE-sided model training data.* |
| Apple [18] | *Proposal 5: For UE side model training data collection, 3GPP only specify the RS used to perform the measurement.* |
| Samsung [20] | *Observation#3: For UE-side model and UE-part of two-sided model, model training*  *Case 1: training at NW-side and model transfer to the UE.*  *Case 2: training by UE-side vendor, e.g., on device or external OTT server*  *The feasibility of Case 1 is strongly tied to the feasibility of model transfer/delivery.*  *Observation#4: For UE-side model and UE-part of two-sided model training by UE-side vendor, proprietary data delivery from UE addresses issues including:*   * *Compatibility on the preferred data format.* * *Auxiliary information needed for model training that may expose proprietary implementation.* * *Data leakage resulting in privacy and security issues.* * *Data ownership issues.*   *Proposal#9: Deprioritize data collection/delivery from UE to entities outside 3GPP network, e.g., OTT server, or to 3GPP network entities other than gNB and LMF.*  *Note: gNB and LMF can collect data based on the same mechanism as network-side model.* |
|  |  |

#### **Background**

During the R18 study item, an LS including the contents of collected training data for different sub use cases were sent to RAN2 [R1-2310681].

RAN2 identified four potential solutions (e.g., 1a, 1b, 2, 3) for data collection for UE-side model training. However, RAN2 didn’t finish the study and no recommendation was agreed.

1. UE collects and directly transfers training data to the Over-The-Top (OTT) server;

1a) OTT (3GPP transparent)

1b) OTT (non-3GPP transparent)

1. UE collects training data and transfers it to Core Network. Core Network transfers the training data to the OTT server.
2. UE collects training data and transfers it to OAM. OAM transfers the needed data to the OTT server.

In RAN#106, the following was agreed for Solution 2 and 3.

From RAN perspective, transfer of data over UP for Solution 2 in SA2 should be studied and provide feedback on fulfillment of visibility and controllability requirements. Transfer of data over UP for Solution 3 can be studied by SA5

RAN has already concluded that transfer of data over CP for Solution 2 and 3 can meet the visibility and controllability RAN requirements, even though some scalability concerns have been raised for high amount of data transfer. RAN is not expecting SA2 and SA5 to study the transfer of data over CP for now.

TSG RAN has also agreed that RAN2 will continue to discuss and study UE sided data collection in WG RAN2 during Rel-19 time frame.

TSG RAN will make a final decision on the way forward taking into account SA inputs when available. TSG RAN will have a checkpoint in Q2, 2025.

**Moderator’s assessment:**

No proposal is suggested for training data collection for UE-sided model. Let’s wait for more inputs.

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Model transfer/delivery

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI [1] | *Proposal 14: For model transfer/delivery Case z4, for two-sided models, consider standardizing both NW-part and UE-part models to help performance monitoring and root cause analysis.*  *Observation 3: Case z1 may be deprioritized as it does not provide clear benefit over Case y; the major difference is the location of their storage. We can start with two options: Case y for 3GPP-transparent scenarios and Case z4 for non-3GPP-transparent scenarios.*  *Proposal 15: Deprioritize Case z1 as it does not provide clear benefit over Case y.*  *Proposal 16: Regarding the study of necessity/benefit of model transfer/delivery Case z4, focus on the option in which the known model structure(s) is standardized by 3GPP.*  *Proposal 17: Regarding standardized known model structure(s) for model transfer/delivery Case z4, if standardizing both NW-part and UE-part models is not feasible, then the study prioritizes the case that UE part of two-sided model is standardized.*  *Proposal 18: Regarding standardized known model structure(s) (if supported) for model transfer/delivery Case z4, study the feasibility of specifying model structures based on the following backbones.*   * *Opt.1: Transformer* * *Opt.2: CNN* |
| ZTE [2] | *Proposal 6: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer study for two-sided model rather than UE-side model.*  *Observation 5: The overall prioritization up to RAN1#118bis is of the following.*   |  |  |  | | --- | --- | --- | |  | *UE-sided model* | *Two-sided model* | | *Case y* |  |  | | *Case z1* |  |  | | *Case z2* | *Deprioritized (RAN1#116bis)* | *Deprioritized (RAN1#118bis)* | | *Case z3* | *Deprioritized (RAN1#116bis)* | *Deprioritized (RAN1#116bis)* | | *Case z4* |  |  | | *Case z5* | *Deprioritized (RAN1#116)* | *Deprioritized (RAN1#116)* |   *Proposal 7: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer z4 for two-sided model.*  *Observation 6: If a proper standardized model backbone with a set of typical hyper-parameters is adopted, the impact on the inference performance is mainly due to the trained model parameters, since the derived model parameters are strongly correlated to the training dataset, which can implicitly reflect the practical channel environment, e.g., a set of cell-specific model parameters.*  *Proposal 8: For the standardization of model structure, Transformer should be adopted as the backbone structure for two-sided model of CSI compression as a starting point.*  *Proposal 9: RAN1 and RAN4 can share the same standardized model structure for two-sided model of CSI compression to avoid the duplicated specification efforts, where the detailed specification and testing can be up to RAN4 study.*  *Proposal 10: For model with Transformer backbone, the following hyper-parameters can be further studied for model structure standardization:*   * *Alternative 1: Typical Transformer backbone*   + *Number of Transformer blocks*   + *Model dimension*   + *Hyper-parameters in the multi-head self-attention modules, at least including number of attention heads and the dimension of each attention head*   + *Hyper-parameters in the feed-forward modules, at least including dimension of latent space, and the activation function* * *Alternative 2: Swin Transformer backbone*   + *Input dimension*   + *Number of Swin-Transformer blocks*   + *Model dimension*   + *Hyper-parameters in the patch partition modules, at least including patch size, and patch resolution*   + *Hyper-parameters in the multi-head self-attention modules, at least including number of attention heads, the dimension of each attention head, window size, QKV bias*   + *Hyper-parameters in the feed-forward modules, at least including dimension of latent space, and the activation function*   *Proposal 11: Regarding model parameter transfer, the detailed mechanism and signalling of model parameter transfer is up to RAN2 discussion.* |
| Huawei [3] | *Observation 2: For model transfer/delivery Case z4, how to align the model structure between NW side and UE side may need further study, e.g., 2 candidates are listed in below:*   * *Candidate 1: Offline alignment between NW side and UE side.*   + *The burden of cross-vendor collaboration still exists.*   + *It causes burden of maintenance/storage of multiple models to different UE vendors at the NW side.* * *Candidate 2: 3GPP specified model structure.*   + *Avoid the burden of cross-vendor collaboration and the burden of maintaining/storing multiple models at NW.*   + *Whether it is possible to achieved agreed-upon model structure at 3GPP level may be questionable.*   + *The common specified model structure may limit the upper bound of the achievable performance of the model.*   *Observation 3: For model transfer/delivery where the model is trained at UE side or neutral site, the necessity of introducing Case z1 as opposed to the implementation manner of Case y is not clear:*   * *Case z1 incurs the burden of offline cross-vendor collaboration, compared to Case y.* * *Case z1 may come with 3GPP NW side burden on model maintenance/storage compared to Case y.* * *Case z1 does not bring benefits compared to Case y.*   *Proposal 6: Regarding the study of model transfer/delivery Case z4,*   * *Partial parameter transfer can be discussed after more progress has been achieved for the specific standardized model structure.* * *Model ID related discussion can be handled by RAN2.*   *Proposal 7: For model transfer/delivery where the model is trained at UE side or neutral site, assume Case y as the baseline.* |
| Spreadtrum [4] | *Proposal 2: From RAN1 perspective, the model transfer/delivery Case z1 should be deprioritized in Rel-19.*  *Proposal 3: For the procedure of model delivery/transfer Case z4, support Alt.B.*  *Proposal 4: Regarding the relationship of model ID, first indication, and second indication for model transfer/delivery Case z4, support Opt.1.* |
| CATT [5] | *Proposal 7: For Alt.A of model transfer case z4, a Step A-0 can be added before Step A-1, in which the NW sends a request to UE on reporting the supported known model structure to NW.*   * *This assumes that Step A-1 is not part of UE capability report.*   *Proposal 8: For Alt.B of model transfer case z4,*   * *If Step B-0 is not considered as part of UE capability report,*    + *If NW indication in Step B-1 is UE-specific signaling, Step B-0 should happen before Step B-1;*   + *If NW indication in Step B-1 is broadcast signaling, Step B-0 should happen after Step B-1;* * *If Step B-0 is considered as part of UE capability report, there is no Step B-0.*   *Proposal 9: Regarding whether/when UE needs the transfer of new parameters, the following case can be considered:*   * *Case 1: UE initiates new parameter transmission, e.g. UE can only store/use one set of parameters, but it moves to a new cell which has cell-specific model to be used.* * *Case 2: NW initiates new parameter transmission, e.g. NW updates its model according to the latest collected data, and would like to transmit the parameters to a UE.*   *Proposal 10: Transfer of partial of the parameters is possible but can reuse the same procedure as full parameters transmission.*  *Proposal 11: In Rel-19, for model transfer case z4, assume the known model structure is standardized model structure.*  *Proposal 12: For model transfer Case z4, the following directions can be considered to align the understanding on supported known model structure between UE and NW:*   * *Direction 1: Standardized reference model + exchange model ID;* * *Direction 2: Standardized reference model structure + exchange model structure ID.*   *Proposal 13: For the relationship between model ID, first indication, and second indication, further study Opt.1 and Opt.3.*   * *Opt.1: Model ID consists of the information of the first indication and the second indication;* * *Opt.3: Model ID is assigned by network and is separated from the first indication and the second indication.* |
| China Telecom [6] | *Proposal 1: The model transfer/delivery Case z1 is deprioritized in Rel-19.*  *Proposal 2: For model delivery/transfer Case z4, prioritize on the study of Alt. b. (Step B-0 ~ Step B-3).*  *Proposal 3: Only single solution is supported to determine the readiness of AI model with the transferred parameters for inference, i.e., either Option 1 (UE sends signaling) or Option 2 (minimum applicable time). And Option 1 is preferred.*  *Proposal 4: Regarding the relationship of model ID, first indication, and second indication for model transfer/delivery Case z4, model ID consists of the information of the first indication and the second indication.* |
| CMCC [7] | *Proposal 7: Model transfer/delivery can have the following usages:*  *1) Model deployment for one-sided model and two-sided model*  *2) Model pairing for two-sided model*  *3) NW-side additional conditions consistency between training and inference*  *Proposal 8: It is suggested to further study model transfer/delivery Case z4, from the following aspects:*   * *How to standardize reference model structure by 3GPP* * *How to exchange model parameters over the air* * *The associated procedure*   *Proposal 9: For the study of model delivery/transfer Case z4, further study the following option 2 for the open format (including the feasibility/specification efforts)*   * *Option 2: Define a new open format within 3GPP (including Using ASN.1 to represent the AI model)*   *Proposal 10: At least after offline engineering at UE side (if needed) or the performance of previous transmitted parameters cannot achieve the performance requirements, transferring new parameters for the known model structure may be needed.*  *Proposal 11: For the case of transfer partial parameters, if the parameters of an AI/ML model consist of multiple partial parameters, then each of the partial parameters should be associated with a same second indication.* |
| vivo [8] | *Proposal 11: Conclude that model transfer in open format of a known model structure at UE (i.e., Case z4) is feasible from device implementation perspective.*  *Observation 4: The burden of model storage would be relieved if the model structure is specified in 3GPP.*  *Observation 5: Proprietary design disclosure may not be a concern if the model structure is specified in 3GPP.*  *Proposal 12: RAN1 can further conclude on feasibility of model parameter update for Case z4 with model structure specified in 3GPP.*  *Proposal 13: For model parameter exchange over air interface, consider using ASN.1 signaling as the starting point. Other open format signaling can be further studied.*  *Proposal 14: For model delivery/transfer Case z4, conclude that it is necessary for UE to report information on whether it needs the transfer of new parameters for a known model structure.*  *Proposal 15: For model delivery/transfer Case z4, conclude that UE sends signaling to network to notify that the AI model with the transferred parameters is ready for activation, after Step A-2 or B-3.*  *Proposal 16: For model delivery/transfer Case z4, conclude that NW could indicate to transmit partially model parameters, in or before Step A-2 or Step B-3.*  *Proposal 17: Model can be referred by the combination of first indication (model structure indication) and second indication (transferred parameter indication).*  *Proposal 18: First indication (model structure indication) is global. Second indication (transferred parameter indication) could be global or local.*  *Proposal 19: The design of first indication (model structure indication) and second indication (transferred parameter indication) should consider partial parameters transfer.* |
| Ericsson [9] | *Proposal 2 RAN1 to prioritize case y for model delivery*  *if the collaboration burden of case y with NW-sided training is deemed infeasible, prioritize case z4 with specified model structure and coefficient precision.* |
| OPPO [10] | *Proposal 8: For model delivery/transfer Case z4, Alt. B is slightly preferred due to smaller signaling overhead.*   * *In Step B-1 and B-2, “Model structure ID” may be used for NW and UE to indicates the candidate list and reports the supported list, respectively.* * *In Step B-3, NW would indicate the complete model ID corresponding to the transferred parameters. This step serves as model identification type B MI-Option 2.* * *Both the model structure ID and the complete model ID are global ID.* * *In the inference stage, NW can assign a local model ID corresponding to the global complete model ID for configuring/indicating the model.*   *Proposal 9: Regarding model transfer/delivery Case z4 for inference,*   * *UE needs the transfer of new parameters for a known model structure when the parameters need to be updated to adapt to a new scenario.* * *UE can report when a model is ready for inference, using a single report for dynamic change of AI/ML capabilities for all kinds of reasons, incl. change of applicable computation power, memory, battery life, and the transferred parameters is ready. No dedicate signaling for reporting the transferred parameters is ready.* * *Deprioritize the study for transfer of partial of the parameters for a known model structure in Rel-19.* |
| Google [11] | *Proposal 3: Support the NW and UE to maintain the same understanding on when the UE can perform data collection.*  *Proposal 4: Support the NW to configure whether the associated ID in different cells indicating the same additional conditions or not.* |
| LG [13] | *Proposal#4. Focus on discussing the key challenges of model transfer such as offline cross-vendor collaboration, model storage requirements, and proprietary design disclosure issues.*  *Proposal#5. Whether/when UE needs the transfer of new parameters for a known model structure can be known by NW via performance monitoring procedure.*  *Proposal#6. Whether/when the AI model with the transferred parameters is ready for inference can be known by NW via defining model application time, i.e. the minimum required time for UE to apply the transferred model, i.e. Option 2.*  *Proposal#7. Model ID should infer both structure information and corresponding parameter set information. Details of model ID, first indication, and second indication should be left to WI phase after clarifying signaling procedures related to such IDs.* |
| NEC [14] | *Observation 8: Supporting model transfer is essential when considering cell/scenario-specific AI/ML deployment which is expected to happen when AI/ML deployment accelerates.*  *Proposal 7: Support Alt. B for model transfer methodology z4.*  *− In Step B-0, UE reports to NW (within UE capability information) that model transfer is supported for which AI/ML features (in Rel-19 only CSI compression use case)*  *− 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*  *Proposal 8: Support RRC signaling for transfer of AI/ML model parameters from gNB to UE.*  *− Discuss further if UE can store the model parameters when it goes into idle/inactive state so that gNB can avoid providing the full model parameters to the UE when UE switches to connected state*  *Proposal 9: For model management purposes, support model structure within the first indication to be identified/associated with a model structure id and the model parameters in the second indication to be identified/associated with a model id value, i.e., Option 2*  *− A transferred AI/ML model is uniquely identified at least within the scope of the cell, where AI/ML model transfer occurs, using both model structure id and model id value*  *Proposal 10: At least to support offline engineering of the model at the UE side, support Option 1: UE sends signaling to network to notify that the AI model with the transferred parameters is ready for activation.* |
| Lenovo [15] | *Observation 3: For the model transfer/delivery Case z4, the UE part model is not always one-by-one mapping with the datasets or model structure.*  *Proposal 4: The supported known model structure, i.e., the first indication, can be indicated in all model identification options, which can be separated from the mode ID.*  *Proposal 5: For model transfer/delivery Case z4, the second indication referring to the transmitted parameters can be assumed as the model ID, i.e., Opt.2, and the relation between the model ID and ID-X need further study.* |
| NVIDIA [16] | *Proposal 4: Conclude that there is a need to consider standardised solutions for transferring/delivering AI/ML models.*  *Proposal 5: Continue to study Cases y, z1 and z4 for transferring/delivering AI/ML models.* |
| Xiaomi [17] | *Observation 1: For the model trained with UE side or neutral site, the need to consider standardised solutions for transferring/delivering AI/ML model(s) is weak.*  *Observation 2: It is beneficial to support that AI models are trained with the network and then delivered/transferred to UE.*  *Observation 3: For Case y with NW side training*   * *Large offline-coordination effort is required;* * *Large time-scale for model update;* * *Potential specification effort on the assistance signalling/procedure for the model transfer/delivery is necessary.*   *Observation 4: For case z4, the following two options are possible for the known model structure alignment between NW and UE*   * *Option 1: Via offline coordination;* * *Option 2: Via 3GPP specification.*   *Observation 5:*   * *For Case z4 with offline known model structure, offline co-ordination effort is required.* * *For Case z4 with specified known model structure, additional specification effort is required. But on the other hand, it could further facilitate the test for RAN4.*   *Proposal 1: Consider standardised solutions for model transfer/delivery at least for the case that AI models are trained on network side.*  *Proposal 2: When the AI models are developed by the network side, prioritize investigating model transfer/delivery solution case z4 with specified known model structure.*  *Proposal 3: Regarding to model transfer/delivery Case z4, the following steps should be considered:*   * *Identify the potential need for the model parameter delivery.* * *Confirm UE’s consent on the model parameter delivery.*   *Proposal 4: Regarding to model transfer/delivery Case z4, two options for model delivery should be considered:*   * *Option 1: NW triggers model delivery and UE confirms.* * *Option 2: UE riggers model delivery.*   *Proposal 5: Regarding to model transfer/delivery Case z4, support UE sending signalling to NW to notify that the model with transferred parameters is ready for activation.*  *Proposal 6: Regarding to model transfer/delivery Case z4, support option 1 to determine the model ID based on the first indication and the second indication.*   * *Option 1: Model ID consists of the information of the first indication and the second indication.* |
| Apple [18] | *Proposal 2: From RAN1 perspective, the model transfer/delivery Case z1 is deprioritized at least for UE-sided model in Rel-19 due to the following reasons:*   * *Risk of proprietary design disclosure* * *Burden of offline cross-vendor collaboration* * *Additional burden on model storage within 3GPP network*   *Proposal 3: For model transfer z4, to determine whether/when UE needs the transfer of new parameters for a known model structure:*   * *NW indicate the model ID and related meta information to the UE before model transfer.* * *UE confirm whether model transfer is required. UE can send negative indication if the model is already transferred before.*   *Proposal 4: For the study of model delivery/transfer Case z4, if inter-vendor training collaboration option 3a is used, to determine the readiness of AI model with the transferred parameters for inference,*   * *UE indicating whether model is ready for inference* |
| Samsung [20] | *Proposal#10: Deprioritize study on Case z1 of 3GPP non-transparent model transfer cases as it requires offline cross-vendor collaboration.*  *Observation#5: For Case z4, model transfer in open format of a known model structure at UE, the exact model structure can be identified between NW and UE through specification.*  *Proposal#11: Study the feasibility and potential benefits of model (parameter) transfer for specified model structure from gNB to UE, i.e., Case z4.*  *Observation#6: For model delivery/transfer Case z4, when model structure is specified, Alt A is feasible, i.e., it is feasible for the UE to report the supported model structure(s) for an AI/ML feature.*  *Proposal#12: For model delivery/transfer Case z4 with specified model structure, further study the necessity of model identification starting from MI-Option4.* |
| ETRI [21] | *Proposal 3: For UE part of two-sided model, further study the following example of MI-Option3 and its feasibility and necessity:*   * *For Case 1 (Offline engineering using delivered/transferred model)*   + *Step A-1: Model is delivered/transferred from the NW/NW-side to UE/UE-side via over-the-air standardized signaling (using model delivery/transfer case z4)*   + *Step A-2: UE part of two-sided model(s) is(are) developed based on at least the above model*   + *Step A-3: UE reports information of its UE part of two-sided model(s) corresponding to the above model to the NW* * *For Case 2 (Direct inference using delivered/transferred model)*   + *Step B-1: Model is delivered/transferred from the NW/NW-side to UE/UE-side via over-the-air standardized signaling (using model delivery/transfer case z4)*   + *Step B-2: UE reports information of its UE part of two-sided model(s) to the NW* * *FFS: How model ID is determined/assigned for each AI/ML model* |
| Fujitsu [22] | *Proposal-5: Deprioritize Case z1 if its benefit over Case y from the location of model storage cannot be justified.*  *Proposal-6: In Case z4, the assumption of “known model structure” is specified by 3GPP.*  *Proposal-7: Regarding model transfer/delivery Case z4, the study on standardized known model structure(s) is for UE part of two-sided model. While the study on standardized known model structure(s) for UE-sided model is suggested to be deprioritized.*  *Proposal-8: For model delivery/transfer Case z4, whether and how to check model performance of the AI model with the transferred parameters for inference are suggested to be studied. The model performance can be checked by:*   * *Option 1: the mechanisms to identify the root cause of performance degradation after activation if the model delivery/transfer is directly used for inference* * *Option 2: the performance assessment before activation for the model delivery/transfer*   *Proposal-9: The transfer of new parameters for a known model structure should take the following issues into consideration:*   * *The first indication referring to the known model structure is the same between NW and UE* * *The second indication referring to model parameters of the same known model structure is different between NW and UE* * *NW makes the final decision of model transfer based on the first and the second indication information reported from UE* * *UE-initiated transfer and NW-initiated transfer*   *Proposal-10: The study on transfer of partial parameters for a known model structure should be deprioritized due to lack of evaluation results and additional workload to inter-vendor collaborations.* |
| Nokia [23] | *Observation 5: The model transfer/delivery Case z4 applies only when the inter-vendor collaboration Option 3a (Direction A) is adopted and requires the model identification MI-Option 3 to be used.*  *Observation 6: The RAN1 related configurations to be considered for the transfer should be flexible enough to allow full or partial model updates, i.e. the CP config might indicate this, and the transfer needs to include UE-vendor specific meta information.*  *Proposal 3: The model transfer/delivery Case z4 configurations for two-sided models, need to support the transfer of a model not ready for inference, including: model structure ID (indication referring to known model structure) and full or partial set of parameter values.*  *Proposal 4: For the model delivery/transfer Case z4, if the model delivery/transfer is directly used for inference, the model can be assumed to be ready for activation from a minimum applicable time after the completion of the transfer of the model structure indication and the model parameters.* |
| Continental Automotive [24] | *Proposal 5: Prioritize Opt.3 (e.g., as separating model ID from indications for greater scalability and adaptability).* |
| AT&T [25] | *Proposal 10: Model transfer/delivery is supported for both UE-sided models and UE-part of two-sided models in Rel-19.*  *Note: Which aspects of model transfer/delivery are supported should be discussed in each sub-use-case.*  *Proposal 11: Regarding the study of necessity/benefit of model transfer/delivery Case z4 for inference, RAN1 focuses on the option with standardized known model structure(s) (i.e., Opt.1)*   * *Note: Offline alignment of the known model structure(s) (i.e., Opt.2) between UE and network is beyond RAN1’s expertise*   *Proposal 12: Regarding the option with standardized known model structure(s) (i.e., Opt.1) for model transfer/delivery Case z4 for inference, the study prioritizes the case that the standardized known model structure(s) of UE-sided model / UE part of two-sided model is standardized.*  *Proposal 13: For the study of model delivery/transfer Case z4, further study the following options for the open format (including the feasibility/specification efforts)*   * *Option 1: Reuse the existing open format(s) that has existed in the AI community (e.g., ONNX)*   + *FFS: which open format(s)* * *Option 2: Define a new open format within 3GPP (including Using ASN.1 to represent the AI model)* * *Option 3: Reuse the mechanism defined in SA2 (interoperability token) for aligning model description format.* |
| Qualcomm [26] | *Proposal 2: Conclude that model transfer/delivery case z4 is deprioritized for UE-sided models in Rel-19 use cases.*  *Proposal 3: Conclude that model transfer/delivery case z4 for inference is deprioritized for two-sided CSI compression use case.* |
| NTT DOCOMO [27] | *Proposal 4: Deprioritize case z1, unless explicit gain of case z1 compared to case y with UE side training is observed.*  *Proposal 5: UE should report the indication that transferred model is ready for inference, when the compiling is necessary for transferred model.*  *Proposal 4: Discuss how many model structures should be standardized for model transfer case z4.*  *Proposal 5: Discuss the approaches to determine which model structures should be standardized. One practical approach is simulation evaluation with the calibration over companies.*  *Proposal 6: Study the pros and cons of using the existing model format (e.g., ONNX) or introducing new 3GPP format.* |

#### **Background**

During the R18 study item, companies have quite divergent views on whether to support AI/ML model transfer/delivery or not and no consensus was achieved.

The outputs of R18 SI on model delivery/transfers are mainly captured in Section 4.3 and Section 7.2.1.4 of TR 38.843 (v2.0.1):

* Six model delivery/transfer cases (i.e., Case y, z1, z2, z3, z4 and z5) are identified and some pros/cons of the cases are also observed/concluded in RAN1 (Section 4.3)
* Eight potential standardized solutions for model transfer/delivery (i.e., Solution 1a, 2a, 3a, 1b, 2b, 3b, 4a and 4b) are identified and the analysis of each potential solution from 4 areas (i.e., A1, A2, A3 and A4) are captured in RAN2 (Section 7.2.1.4)

|  |  |  |  |
| --- | --- | --- | --- |
| **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. | | | |

During the R19 discussions, some conclusion/agreements were achieved to deprioritize R19 study on some cases. The current status is summarized in the following table:

|  |  |  |
| --- | --- | --- |
| Model delivery/transfer | UE-sided model | Two-sided model |
| Case y |  |  |
| Case z1 |  |  |
| Case z2 | Deprioritized | Deprioritized |
| Case z3 | Deprioritized | Deprioritized |
| Case z4 |  |  |
| Case z5 | Deprioritized | Deprioritized |

## 1st round discussion

#### **Proposal 4.1**

Regarding how to identify the “known” structure(s) for Case z4, there are different options:

* The known structure(s) is specified in 3GPP (same as Option 3 of CSI compression)
* The known structure(s) is identified via offline coordination between vendors
* …

The different options have their own pros and cons.

**Proposed 4.1**

**Observation**

**Regarding the different options to align the same understanding between UE and network on the “known model structure(s)” for the model transfer/delivery z4, RAN1 has the following observations:**

* **Opt.1: The known model structure(s) is specified by 3GPP**
  + **It can alleviate the burden of cross-vendor collaboration on the alignment of model structure(s)**
  + **Large standardization efforts on specifying model structure(s) in 3GPP**
  + **A limited number of specified model structure(s)**
  + **UE’s implementation needs to be compatible to a limited number of specified model structure(s)**
* **Opt.2: Offline alignment between NW and UE, i.e., the known model structure is not specified by 3GPP** 
  + **It can reduce the standardization effort**
  + **There is burden of cross-vendor collaboration on the alignment of model structure(s)**
  + **It may lead to burden of maintenance/storage of multiple models to different UE vendors at the NW side**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | We have these discussions in CSI compression. It is hard to see the need of repeating these in this sub-agenda item. |
| Panasonic | To have use case specific discussion can be more suitable as the choice of options above would not be generic. |
| Fujitsu | OK. |
| Huawei, HiSilicon | OK. |
| TCL | Agree |
| Xiaomi | Support |
| QC | I think the understanding from Rel-18 study is Option 2. The term “known model structure” was agreed during Rel-18 study, and the model structure standardization discussion came up in Rel-19 study. So, I prefer not to change the definition, but instead, re-word the discussion into whether we want to limit the support of z4 into specified models or other models.  Agree with Nokia and Panasonic that there is no need for duplicate efforts as three are related discussions in the CSI compression agenda item. |
| ETRI | We agree. |
| ZTE | Ok to have the discussion either in CSI compression agenda or in framework agenda.  In addition, we suggest to delete the second bullet “Large standardization efforts on specifying model structure(s) in 3GPP” of opt.1 since in the end RAN1 may directly reuse the model structure specified in RAN4. The spec effort in RAN1 may not be large. |
| CMCC | Support. |
| CATT, CICTCI | OK to make it clear. Though our preference is Opt.1, it is indeed two different options in companies opinions, and thus we are OK to spell it out. |
| LG | Fine with the direction. |
| Spreadtrum | Fine with the direction. |
| Futurewei | Agree with the observation. |
| Mod | @QC: I failed to get your point. Could you show me some hints/evidence for your first part of comments. |
|  |  |

#### **Proposal 4.2**

The following proposal is suggested based on the following reasons:

* Offline alignment of the known model structure(s) between UE and NW is out of RAN1 expertise
* Reduce the workload

**Proposal 4.2**

**Agreement**

**Regarding the study of necessity/benefit of model transfer/delivery Case z4, RAN1 focuses on the option with standardized known model structure(s) (i.e., Opt.1).**

* **~~Note: Offline alignment of the known model structure(s) (i.e., Opt.2) between UE and network is beyond RAN1’s scope~~**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Ok with the main bullet. We do not have to discussion Opt.2. |
| Panasonic | Share the view with Nokia. |
| Fujitsu | OK. |
| Huawei, HiSilicon | OK. BTW we think Option 5b in 9.1.4.1 has already been deprioritized? |
| TCL | Agree |
| Xiaomi | Support |
| QC | Question for clarification: Is the proposal only for two-sided CSI compression or for future use cases? For future use cases, we cannot agree on the proposal.  As Huawei noted, Option 5b is already deprioritized, so we don’t need this agreement if the intention is for two-sided CSI compression. |
| ETRI | This is not for use-case specific, so we prefer not to urgently decide on this. |
| ZTE | Ok to keep the main bullet and remove the note. |
| Samsung | Ok to capture the main bullet. |
| CMCC | Support. |
| CATT, CICTCI | Support. |
| Ericsson | Share the view to only keep the main bullet. |
| LG | This is not critical issue as FL can sort out what to discuss in RAN1. Also, this is not for agreement, but for a guidance for future discussions. |
| Futurewei | Agree. |
| Mod | The proposal is updated according to the suggestions from several companies |

#### **Proposal 4.3**

**Proposed 4.3**

**Agreement**

**Regarding the option with standardized known model structure(s) (i.e., Opt.1) for model transfer/delivery Case z4 for two-sided model, the study prioritizes the standardized known model structure(s) of ~~UE-sided model /~~ UE part of two-sided model.**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | OK. |
| Panasonic | OK. |
| Fujitsu | The motivation to study on standardized known model structure(s) for UE-sided model is unclear, it brings the restriction to UE’s differential design and has impact on the performance gain of UE-sided model for reusing or being compatible with standardized model structure(s). |
| Huawei, HiSilicon | Disagree with the UE-sided model part. Given the time left for R19 and the limited progress on specifying model structure, finishing the standardization of two-sided model may even be challenging. We should focus on two-sided model and disregard one-sided model.  **Regarding the option with standardized known model structure(s) (i.e., Opt.1) for model transfer/delivery Case z4, the study prioritizes the standardized known model structure(s) of ~~UE-sided model /~~ UE part of two-sided model.** |
| TCL | Agree |
| Xiaomi | Support |
| QC | Question for clarification: Is the proposal only for two-sided CSI compression or for future use cases? For future use cases, we cannot agree on the proposal.  As Huawei noted, Option 5b is already deprioritized, so we don’t need this agreement if the intention is for two-sided CSI compression. |
| ETRI | Since the updated WID of this agenda is for two-sided model, discussion on UE-sided model is not needed, in our view. |
| ZTE | The relationship between proposal 4.2 and 4.3 should be clarified first. It seems the only difference is that proposal 4.3 mentions UE sided model. However, we believe that model transfer is only useful for the two-sided model. Thus, we proposal to merge proposal 4.2 and 4.3 and avoid mentioning UE sided model. |
| Samsung | Support |
| CMCC | Support. |
| CATT, CICTCI | Support. |
| QC2 | Is this for two-sided CSI compression or future two-sided models?  For two-sided CSI compression, z4 can be deprioritized.  For future use cases, this is premature. |
| Ericsson | OK |
| LG | Unclear on the meaning of ‘prioritize’ in this study. Is this for prioritization for future WI phase? |
| Spreadtrum | Agree with HW. R19 general framework should focus on two-sided model, then “UE sided model” should be deleted. |
| Futurewei | Agree with Huawei’s edit. |
| Mod | The proposal is updated |
|  |  |

#### **Proposal 4.4**

**Proposal 4.4**

**Agreement**

**For model delivery/transfer Case z4 for the inference for two-side model use case(s) (if applicable), the following options are identified for the requesting/triggering the procedure of new parameters transfer from NW to UE for a known model structure**

* **Option 1: UE requests the procedure for new parameters transfer for a known model structure**
* **Option 2: NW triggers the procedure for new parameters transfer for a known model structure**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Model transfer for direction B should be deprioritized, as that is not having good feasibility studies and somewhat not considered in CSI compression. That means RAN1 should not discuss NW to UE model transfer.  When model transfer happens for Direction A, focus shall be NW-side to UE-side, and hopefully RAN2/SA will discuss detailed solutions for that. |
| Panasonic | Support Nokia view. |
| Fujitsu | Support. |
| Huawei, HiSilicon | 1) Case z4 is model transfer from NW to UE (not UE side).  2) As the model transfer direction of Case z4 (i.e., Option 3b/3a-1 in 9.1.4.1) is from NW to UE, why a UE can initiate new parameter transfer (Option 1)? |
| TCL | Agree |
| Xiaomi | Support.  To addiress Huawei’s 2nd concern, maybe changing the word “initiates” to “trigger”/ “request” would be better. |
| QC | Share similar view as Nokia |
| ETRI | This seems on details and may can be discussed under each specific use cases. |
| ZTE | Ok to further study. |
| Samsung | Ok |
| CMCC | Support. And rewording the “initiates” to “request” in the Option 1 is more appropriate. |
| CATT, CICTCI | Support. |
| LG | This is currently under discussion in RAN2. |
| Spreadtrum | Ok to further study. |
| Futurewei | Agree in general. |
| Mod | The proposal is updated based on the comments from several companies  Another controversial point is UE/NW vs UE-side/NW-side. I checked the TR, Huawei’s comment seem correct: UE/NW is used for Case z4 in the TR. |

#### **Proposal 4.5**

In order to assess the feasibility/benefit/spec impact of Case z4 (e.g., for Opt.1), the specification efforts on the open format should also be considered. Thus, based on the tdocs and the discussion of the last meeting(s) the following proposal is suggested for discussion:

**Proposal 4.5**

**Agreement**

**For the study of model delivery/transfer Case z4, if the known structured model is specified in 3GPP, ~~further study the following options for~~ support the following for the open format ~~with down-selection (including the feasibility/specification efforts)~~**

* **~~Option 1: Reuse the existing open format(s) that has existed in the AI community (e.g., ONNX)~~**
  + **~~FFS: which open format(s)~~**
* **~~Option 2:~~ Define a new open format within 3GPP (including Using ASN.1 to represent the AI model)**
* **~~Option 3: Reuse the mechanism defined in SA2 (interoperability token) for aligning model description format.~~**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | As Case Z4 details shall be built on top of the specified model structures, the first step is to identify model structures for two-sided CSI compression (this is already ongoing study). We think that Option 2-like solution (i.e., **a new delivery format for 3GPP specified model structures)** needs to be specified for model transfer. |
| Panasonic | This conclusion can be use case specific. |
| Huawei, HiSilicon | As we already have the model structure specified, why do we still need Option 1 and Option 3 (which are used mainly to describe model structure)? |
| TCL | Agree. |
| Xiaomi | Support |
| QC | OK. We prefer this discussion to be more generic and future-proof, not just focusing on CSI compression. |
| ETRI | We agree |
| ZTE | Same question as Huawei. |
| CATT, CICTCI | OK… But in our understanding, a language of describing model structure (i.e. model description format) is mainly used to express freedom structures. On the other hand, if model structure is specified, we do not need pursue model description format (Option 1 or 3), we just need to find a way to document the specified ones. |
| Ericsson | Same as HW |
| LG | What aspect we need to study each option? |
| Futurewei | Agree in general. |
| Mod | The proposal is updated |
|  |  |
|  |  |

#### **Proposal 4.6**

**Proposed 4.6**

**Agreement**

**For the study of model delivery/transfer Case z4, if the model delivery/transfer is used for training/re-training at UE-side (e.g., UE-side OTT server), the following candidate solution is identified**

* **UE sends signaling to network to notify that the UE supports the AI/ML operations compatible to the AI model with the transferred parameters**
* **FFS: details of signaling (e.g., UE capability signaling)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Can be considered, if needed, in the work item stage for CSI compression (if Case Z4 is supported) |
| Panasonic | Support Nokia view. |
| Fujitsu | OK. |
| Huawei, HiSilicon | Not clear about the intention of this proposal. UE indicating its support to the to-be-delivered model structure has been agreed previously (as highlighted).  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 |
| TCL | Agree |
| Xiaomi | Similar comment with Huawei, the intention of the proposal can be further clarified |
| QC | Also agree that the proposal is not needed. |
| ZTE | Ok to study this. But it seems this it not an urgent issue for this study item, which actually can be discussed in normative phase. |
| CMCC | We want to clarify that the signaling to network happens after one operation of model transfer, it is used to indicate the AI model with the transferred parameters is ready for inference. |
| CATT, CICTCI | Support.  To HW and xiaomi, we understand that this proposal is to indicate that the UE already deployed the compiled model (with transferred model parameters) from its server and ready for inference. Not to indicate its support of model structure. |
| Ericsson | This proposal seems premature. |
| LG | As other companies, further clarification is needed on the intention of the proposal. |
| Futurewei | Agree with Ericsson that this proposal seems premature. This could be done by UE capability report. |
| Mod | The proposal is updated   * Huawei/xiaomi: please see the comment from CATT * Nokia/Panasonic, …: The intention is to figure out the whole picture/the main compnents of the model transfer/delivery mechanism. The detailed design will not be discussed (e.g., details of the signaling) * FW: FFS part is added |
|  |  |

#### **Proposal 4.7**

**Proposed 4.7**

**Agreement**

**Regarding the standardized known model structure(s) (if supported) for model transfer/delivery Case z4 for inference, consider the following backbone as a starting point:**

* **Opt.1: Transformer**
* **Opt.2: CNN**
* **Opt.3: LSTM**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | There is an agreement already in CSI compression to use Option 1 as the starting point. |
| Panasonic | To be discussed in CSI compression. |
| Huawei, HiSilicon | Opt1 has been agreed in 9.1.4.1 at #119 |
| TCL | Support Opt. 1. |
| QC | As mentioned by Nokia and Huawei, the proposal is not needed. |
| ZTE | Similar view as Nokia, Huawei and Qualocmm. |
| CMCC | It is more like use case specific discussion. |
| CATT, CICTCI | Somehow this proposal may have to be use case specific.  Of course Opt.1 is already agreed in AI 9.1.4.1. |
| Ericsson | Similar view as Nokia, Huawei,ZTE and Qualocmm |
| LG | Specific backbone is use-case specific. |
| Spreadtrum | Which model backbone to use should be use-case specific. |
| Futurewei | Agree with Opt 1 and Opt 2. |
|  |  |

#### **Proposal 4.8 (Placeholder)**

**Proposal 4.8 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Issues related to terminology alignment between RAN and SA

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Panasonic [19] | *Proposal 1: to modify RAN1 TR below and to remove SA TR note.*  *Functionality-based lifecycle management: Signaling procedure where network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signalling (e.g., RRC, MAC-CE, DCI); operates based on, at least, one configuration of AI/ML-enabled Feature/FG or specific configurations of an AI/ML-enabled Feature / Feature Group. Training, testing and maintenance phases can be standardized or implementation-specific.*  *Proposal 2: to remove following part in SA TR as following.*  *~~The definition of Federated Learning provided by RAN WG1 appears to only apply to Horizontal Federated Learning, as the phrase "each performing local model training using local data samples" implies that the data samples at individual nodes are distinct. The key difference between Horizontal Federated Learning and Vertical Federated Learning lies in the characteristics of the local datasets:~~*  *~~- Horizontal Federated Learning: Local datasets have the same features but different samples.~~*  *~~- Vertical Federated Learning: local data sets have different features but share same samples.~~*  *Proposal 3: ML model description is modified in RAN1 TR as follow.*  *ML model: A mathematical algorithm that applies ML techniques to generate a set of outputs based on a set of inputs. It may include metadata which consists of, e.g., information related to the model, and applicable runtime context.*  *Proposal 4: ML model training is modified in RAN1 TR as follow.*  *ML model training: A process to train an ML Model by learning the input/output relationship in a data driven manner and obtain the trained ML Model for e.g. inference.*  *Proposal 5: "ML model parameter update" and "ML model update" is not merged to "ML model re-training" in SA TR. RAN1 TR keeps "ML model parameter update" and "ML model update".*  *Proposal 6: ML model testing is modified in RAN1 TR as follow.*  *ML model testing: A process of evaluating the performance of an ML model using test data different from data used for model training and validation.*  *Proposal 7: ML model inference is modified in RAN1 TR as follow.*  *ML model inference: A process of running a set of inputs through a trained ML model to produce a set of outputs.*  *Proposal 8: ML model activation and deactivation are modified in RAN1 TR as follow.*  *ML model activation: A process to enable an ML model for inference.*  *ML model deactivation: A process to disable an ML model for inference.* |
|  |  |

#### **Background**

In Rel-18 and Rel-19, most working groups in TSG SA, TSG CT and TSG RAN have already performed SIs and/or have WIs relating to the AI/ML topic. Meanwhile, due to various factors, there are some potential misalignments/inconsistencies in the definition and use of AI/ML LCM across 3GPP WGs.

In order to address the afore-mentioned inconsistencies, SA is working on TR 22.850 for better alignment within 3GPP across different AI/ML related initiatives.

The tdoc R1-2500815 identified some potential issues of the current TR 22.850. From moderator’s perspective, this tdoc raises some interesting points that are worthy of being discussed.

## 1st round discussion

Regarding the proposals in R1-2500815, moderator’s feeling is

* Some proposal(s) seems essential as they correct some misunderstanding in different groups.
* Some proposal(s) seems “better to have”
* Some proposal(s) may be controversial

As this meeting is the first time to discuss these kinds of things, moderator lists all the proposals from R1-2500815 to check companies’ views.

* If some modification(s) for SA TR 22.850 is agreed, we can send a LS to SA
* If some modification(s) for TR 38.843 is agreed, we can prepare some pCR in future meeting(s).

#### **Proposal 5.1**

For Proposal 1 of R1-2500815

* Moderator shares the same view that Note 1 in Section 6.2.1.8 of TR 22.850 does not correctly capture RAN1 study
* The new added part seems not essential

**Proposed 5.1**

**Agreement**

**From RAN1 perspective, it is suggested to remove ~~Remove~~ the part highlighted by Yellow from Section 6.2.1.8 of TR 22.850 as below**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6.2.1.8 Analysis on ML model lifecycle management SA WG5 describes the ML model lifecycle in clause 4a.0 of TS 28.105 [9], and ML model lifecycle management capabilities for ML model training, ML model testing, ML inference emulation, ML model deployment and ML inference in clause 6.1 of TS 28.105 [9]. The terms ‘ML model-based lifecycle management’, ‘ML-enabled functionality’, and ‘Functionality-based lifecycle management’ have been defined by RAN1, as illustrated in Table 6.2.1.8-x.  Editor's note: Further analysis may be needed, e.g. to determine whether a unified definition can be derived.  Table 6.2.1.8-x: Definitions of ML model-based lifecycle management, ML-enabled functionality and Functionality-based lifecycle management as defined across 3GPP WGs.   |  |  | | --- | --- | | **TSG (TS/TR)** | **ML model lifecycle management / Functionality-based lifecycle management** | | 3GPP RAN1 TR 38.843 [7] | **ML model-based lifecycle management:** Operates based on identified logical models, where a model may be associated with specific configurations/conditions associated with UE capability of an AI/ML-enabled Feature / Feature Group and additional conditions (e.g., scenarios, sites, and datasets) as determined/identified between UE-side and NW-side. The models are identified at the Network, and Network/UE may activate/deactivate/select/switch individual AI/ML models via model ID.  **(ML-enabled) Functionality:** An AI/ML-enabled Feature/Feature Group enabled by configuration(s), where configuration(s) is(are) supported based on conditions indicated by UE capability.  **Functionality-based lifecycle management:** Signaling procedure where network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signalling (e.g., RRC, MAC-CE, DCI); operates based on, at least, one configuration of AI/ML-enabled Feature/FG or specific configurations of an AI/ML-enabled Feature / Feature Group. |   The following unified definition for ‘ML model lifecycle management’ is proposed:  **ML model lifecycle management:** The management capabilities allowing a consumer to manage different phases of the ML model lifecycle as defined in clause 6.2.1.7.  The following definition for ‘Functionality-based lifecycle management’ is proposed for adoption by all 3GPP RAN Working Groups:  **Functionality-based lifecycle management:** Signaling procedure where network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signalling (e.g., RRC, MAC-CE, DCI); operates based on, at least, one configuration of AI/ML-enabled Feature / Feature Group or specific configurations of an AI/ML-enabled Feature/FG.  NOTE 2: Applicability of Functionality-based lifecycle management definition to/in TSG SA WGs is optional. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | Looked ok.  However, RAN1 shall first discuss whether there are provisions to update a TR prepared by SA. Better to suggest changes to SA via a LS than agreeing to CRs in generic form. |
| Panasonic | Support (as we propose it). No intention of official CR. It can be LS or just to ask somebody in SA can be also ok. |
| Huawei, HiSilicon | RAN1 can send LS reply to clarify the definitions. But it seems not urgent; better to leave some time to offline check TR 22.850. |
| TCL | Agree. |
| ZTE | We also suggest to leave more time for companies to thoroughly check the whole TR22.850 and come back in next meeting with a whole TP proposal. |
| Samsung | Support ZTE’s proposal. |
| CATT, CICTCI | Agree. |
| LG | To my understanding, we cannot directly revise TR22.850. |
| Mod | The proposal is updated  @all companies: If RAN1 made some agreements, we will send a LS including the corresponding agreements and ask SA to consider it. RAN1 will NOT provide TP directly. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.2**

**Proposed 5.2**

**Agreement**

**From RAN1 perspective, it is suggested to remove ~~Remove~~ the part highlighted by Yellow from Section 6.2.2 of TR 22.850 as below**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6.2.2 Analysis on Federated Learning The term 'Horizontal Federated Learning' and ‘Vertical Federated Learning' have been defined in SA WG2 and RAN WG1 defines 'Federated Learning', as illustrated in Table 6.2.2-1.  Editor's note: Further analysis may be needed, e.g. to determine whether a unified definition can be derived.  Table 6.2.2-1: Definition of Federated Learning as defined across 3GPP WGs   |  |  | | --- | --- | | TSG (TS/TR) | Federated Learning | | SA WG2 TR 23.700-84 [4] | *Horizontal Federated Learning*: A federated learning technique without exchanging/sharing local data set, wherein the local data set in different FL clients for local model training have the same feature space for different samples (e.g. UE IDs). | | SA WG2 TR 23.700-84 [4] | *Vertical Federated Learning*: A federated learning technique without exchanging/sharing local data set, wherein the local data set in different VFL Participant for local model training have different feature spaces for the same samples (e.g. UE IDs). | | RAN WG1 TR 38.843 [3] | *Federated Learning*: A machine learning technique that trains an AI/ML model across multiple decentralized edge nodes (e.g. UEs, gNBs) each performing local model training using local data samples. The technique requires multiple interactions of the model, but no exchange of local data samples. |   The terms "distributed learning" and "federated learning" are often used together as "distributed/federated learning" in SA1 TS 22.261. "Distributed learning" typically refers to a broader set of learning techniques including "federated learning". Although the two terms are related, they are not identical and should be used appropriately based on the context.  The following unified definition for ‘Horizontal Federated Learning’ is proposed:  **Horizontal Federated Learning:** A federated learning technique without exchanging/sharing local data set, wherein the local data set in different HFL clients for local model training have the same feature space for different samples.  The following unified definition for ‘Vertical Federated Learning’ is proposed:  **Vertical Federated Learning:** A federated learning technique without exchanging/sharing local data set and local ML models, wherein the local data set in different VFL clients for local model training have different feature spaces for the same samples. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Nokia | OK. Same comment as before. |
| Panasonic | Support (as we propose it). |
| Huawei, HiSilicon | RAN1 can send LS reply to clarify the definitions. But it seems not urgent; better to leave some time to offline check TR 22.850. |
| TCL | Agree |
| ZTE | We also suggest to leave more time for companies to thoroughly check the whole TR22.850 and come back in next meeting with a whole TP proposal. |
| CATT, CICTCI | Agree |
| LG | Same comment as before |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.3**

**Proposed 5.3**

**Agreement**

**Use the following definition from TR 22.850 to replace the definition of AI/ML model in TR 38.843**

|  |
| --- |
| **ML model:** A mathematical algorithm that applies ML techniques to generate a set of outputs based on a set of inputs. It may include metadata which consists of, e.g., information related to the model, and applicable runtime context. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | The current definition of AI/ML model in TR 38.843 is as below:   |  | | --- | | **AI/ML Model:** A data driven algorithm that applies AI/ML techniques to generate a set of outputs based on a set of inputs. | |
| Nokia | This is not a critical change. |
| Panasonic | Support. If the question is critical or not, I agree not critical. |
| Huawei, HiSilicon | No need to change TR38.843. It is natural that different WGs have different usages and different emphasis of the definition on AI/ML. |
| TCL | Agree |
| CATT, CICTCI | Hesitate to support, since metadata and runtime context are rarely discussed in RAN1, or captured in TR 38.843. |
| LG | First of all, do we have plan to update TR38.843 in Rel-19 or write a new TR?  For this specific proposal, we also think that this correction is not necessary. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.4**

**Proposed 5.4**

**Agreement**

**Use the following definition from TR 22.850 to replace the definition of AI/ML model training in TR 38.843**

|  |
| --- |
| **ML model training:** A process to train an ML Model by learning the input/output relationship in a data driven manner and obtain the trained ML Model for e.g. inference. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | The current definition of AI/ML model training in TR 38.843 is as below:   |  | | --- | | **AI/ML model training:** A process to train an AI/ML Model [by learning the input/output relationship] in a data driven manner and obtain the trained AI/ML Model for inference. | |
| Nokia | This is not a critical change. |
| Panasonic | Support. If the question is critical or not, I agree not critical. |
| Huawei, HiSilicon | OK to align as there is only editorial difference. But RAN1 term is “AI/ML” instead of “ML” |
| TCL | Agree |
| CATT, CICTCI | RAN1 usually use AI/ML rather than ML. The effort of updating the whole TR seems huge and not meaningful. |
| LG | Not needed. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.5**

**Proposed 5.5**

**Agreement**

**Use the following definition from TR 22.850 to replace the definition of AI/ML model testing in TR 38.843**

|  |
| --- |
| **ML model testing:** A process of evaluating the performance of an ML model using test data different from data used for model training and validation. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | The current definition of AI/ML model testing in TR 38.843 is as below:   |  | | --- | | **AI/ML model testing:** A subprocess of training, to evaluate the performance of a final AI/ML model using a dataset different from one used for model training and validation. Differently from AI/ML model validation, testing does not assume subsequent tuning of the model. | |
| Nokia | This is not a critical change. |
| Panasonic | Support. If the question is critical or not, I agree not critical. |
| Huawei, HiSilicon | No need to change TR38.843. It is natural that different WGs have different usages and different emphasis of the definition on testing. |
| TCL | Agree |
| CATT, CICTCI | Somehow we think RAN1’s definition provide more critical information than SA’s. |
| LG | Not needed. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.6**

**Proposed 5.6**

**Agreement**

**Use the following definition from TR 22.850 to replace the definition of AI/ML model inference in TR 38.843**

|  |
| --- |
| **ML model inference:** A process of running a set of inputs through a trained ML model to produce a set of outputs. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | The current definition of AI/ML model inferece in TR 38.843 is as below:   |  | | --- | | **AI/ML model Inference:**  A process of using a trained AI/ML model to produce a set of outputs based on a set of inputs. |   The two definitions seem the same although the ordering of some words are different |
| Nokia | This is not a critical change. |
| Panasonic | Support. If the question is critical or not, I agree not critical. |
| Huawei, HiSilicon | OK to align as there is only editorial difference. But RAN1 term is “AI/ML” instead of “ML” |
| TCL | Agree |
| CATT, CICTCI | They are almost identical already. |
| LG | Not needed |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.7**

**Proposed 5.7**

**Agreement**

**Use the following definition from TR 22.850 to replace the definition of model activation/deactivation in TR 38.843**

|  |
| --- |
| **ML model activation:** A process to enable an ML model for inference.  **ML model deactivation:** A process to disable an ML model for inference. |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | The current definitions of model activation/deactivation in TR 38.843 are as below:   |  | | --- | | **Model activation:** enable an AI/ML model for a specific AI/ML-enabled feature.  **Model deactivation:** disable an AI/ML model for a specific AI/ML-enabled feature. | |
| Nokia | This is not a critical change. |
| Panasonic | Support. If the question is critical or not, I agree not critical. |
| Huawei, HiSilicon | No need to change TR38.843. It is natural that different WGs have different usages and different emphasis of the definition on activation/deactivation. |
| TCL | Agree |
| CATT, CICTCI | RAN1’s definition focuses on ‘a specific AI/ML-enabled feature’ and seems providing more RAN1 background understanding than SA. |
| LG | Not needed |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 5.8 (Placeholder)**

**Proposal 5.8 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | If some modification(s) for SA TR 22.850 is agreed, we can send a LS to SA |
| Nokia | OK |
|  |  |
|  |  |
|  |  |
|  |  |

# Others

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| NEC [14] | *Observation 9: It is important to discuss how UE can indicate its internal restrictions to activate or run an AI/ML model/functionality to the network for optimal AI/ML operation.*  *Observation 10: Reporting of UE’s internal conditions such as memory size, battery level and other detailed hardware limitations to gNB for AI/ML operation may lead to UE’s proprietary information disclosure and may be hard for network to determine AI/ML applicability for a UE based on the provided information.*  *Proposal 11: Specify UE indication to network about its inability to run a configured/activated AI/ML model/functionality due to UE’s internal condition along with a relevant cause value for the failure.* |
| NVIDIA [16] | *Observation 1: Deterministic, physics-based modelling for wireless propagation, especially ray tracing, are essential for studying, evaluating, and developing AI/ML models in 5G-Advanced toward 6G.* |
|  |  |
|  |  |

**Moderator’s assessment:** No proposal or issue recommended for discussion

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Summary of discussion

## Proposals for Wednesday’s online session

**Proposal 4.4**

**Agreement**

**For model delivery/transfer Case z4 for the inference for two-side model use case(s) (if applicable), the following options are identified for the requesting/triggering the procedure of new parameters transfer from NW to UE for a known model structure**

* **Option 1: UE requests the procedure for new parameters transfer for a known model structure**
* **Option 2: NW triggers the procedure for new parameters transfer for a known model structure**

**Proposal 4.5**

**Agreement**

**For the study of model delivery/transfer Case z4, if the known structured model is specified in 3GPP, ~~further study the following options for~~ at least consider the following for the open format ~~with down-selection (including the feasibility/specification efforts)~~**

* **~~Option 1: Reuse the existing open format(s) that has existed in the AI community (e.g., ONNX)~~**
  + **~~FFS: which open format(s)~~**
* **~~Option 2:~~ Define a new open format within 3GPP (including Using ASN.1 to represent the AI model)**
* **~~Option 3: Reuse the mechanism defined in SA2 (interoperability token) for aligning model description format.~~**

# Appendix A: Agreements

## RAN1#116

Agreement

* 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

* 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

**Conclusion**

From RAN1 perspective, the model transfer/delivery Case z5 is deprioritized for Rel-19.

**Conclusion**

RAN1 has no consensus to reply the SA5 LS (R1-2400035)

## RAN1#116bis

Conclusion

From RAN1 perspective, the model transfer/delivery Case z2 is deprioritized at least for UE-sided model in Rel-19 due to the following reasons:

* Risk of proprietary design disclosure
* Burden of offline cross-vendor collaboration

Conclusion

From RAN1 perspective, the model transfer/delivery Case z3 is deprioritized for Rel-19 due to the following reasons (compared to Case y):

* No much benefit compared to Case y
* Risk of proprietary design disclosure
* Large burden of offline cross-vendor collaboration
* Additional burden on model storage within in 3GPP network

Conclusion

* It is clarified that MI-Option 4 refers to the Option 1 of CSI compression
  + Option 1: Fully standardized reference model (structure + parameters)

Agreement

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)

* 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
* B: UE(s) collects the data corresponding to the associated ID(s)
* C: AI/ML models are developed (e.g., trained, updated) at UE side based on the collected data corresponding to the associated ID(s).
* 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
* 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.

## RAN1#117

Working Assumption

Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent at least within a cell

* FFS: whether/how UE assumption can be applicable for multiple cells (including the feasibility study)

**Agreement**

**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**

**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.**

## RAN1#118

**Conclusion**

* **From RAN1 perspective, model identification is at least applicable to some of inter-vendor training collaboration option(s) of CSI compression using two-sided model (if supported)**

**Conclusion**

**The model identification procedure dedicated to MI-Option5 is not pursued for Rel-19 normative work.**

**Conclusion**

**The model identification procedure dedicated to MI-Option2 for one-sided model is not pursued for Rel-19 normative work.**

**Agreement**

**Confirm the following Working assumption.**

**Working Assumption**

**Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent at least within a cell**

* **FFS: whether/how UE assumption can be applicable for multiple cells (including the feasibility study)**

**Agreement**

**From RAN1 perspective, the “known model structure(s)” of the model transfer/delivery Case z4 at least include known information on the following aspects**

* **Model type/backbone (e.g., Transformer, CNN and so on)**
* **In case model type is a neural network** 
  + **Number of layers**
  + **Layer types/structure (e.g., full connected, activation layer and so on)**
  + **Layer size (e.g., the number of parameters of a layer)**
  + **Connection between different layers**
* **model input/output related information**

**Conclusion**

**From RAN1 perspective, model transfer is needed at least for some (e.g., Option 3b) of inter-vendor training collaboration option(s) of CSI compression using two-sided model (if supported)**

Agreement

RAN1 is recommending extending the study of the Model identification, and Model transfer/Model delivery based on RAN1 understanding the study is not completed.

Agreement

Adopt the TP1, TP2, TP3 and TP4 in Section 2 of R1-2407520 in principle.

## RAN1#118bis

Agreement

Regarding model transfer/delivery Case z4 for inference, further study the following aspects:

* Whether/when UE needs the transfer of new parameters for a known model structure.
* Whether/when the AI model with the transferred parameters is ready for inference.
* The transfer of partial of the parameters for a known model structure.

Agreement

Regarding the study of model transfer/delivery Case z4, for a given known model structure, network can transmit the following information along with the parameters:

* a first indication referring to the known model structure at least for the case there are multiple known model structures
* a second indication referring to the transmitted parameters
* FFS: relationship of model ID, first indication, and second indication
* FFS: transfer partial parameters

Conclusion

From RAN1 perspective, the model transfer/delivery Case z2 is deprioritized for two-sided model in Rel-19.

Agreement

Regarding the study of MI-Option2 (i.e., model identification with dataset transfer) for the two-sided model, ID (denoted as ID-X) can be transmitted from network/network-side to UE/UE-side for the dataset.

* Note: The notation “ID-X” is used for discussion purpose

## RAN1#119

Agreement

For study of MI-Option2 (i.e., model identification with dataset transfer) for the two-sided model,

* ID-X can be used for pairing the UE-part and the NW-part of a two-sided model
* FFS: other information needed for pairing

Conclusion

Regarding MI-Option2 (i.e., model identification with dataset transfer) for the two-sided model, from RAN1 perspective, how to construct the dataset, including whether a dataset constructed from one cell or from multiple cells is up to network implementation.

Agreement

Regarding the relationship of model ID, first indication, and second indication for model transfer/delivery Case z4, further study the following options:

* Opt.1: model ID consists of the information of the first indication and the second indication
  + E.g., model ID is a combination of the first and second indications
* Opt.2: The second indication is assumed as the model ID
* Opt.3: Model ID is assigned by network and is separated from the first indication and the second indication

Conclusion

For the study of model delivery/transfer Case z4, if the model delivery/transfer is directly used for inference, the following options are identified as the candidate solutions to determine the readiness of AI model with the transferred parameters for inference (either or combination of the following options)

* Option 1: UE sends signaling to network to notify that the AI model with the transferred parameters is ready for activation
* Option 2: The AI model with the transferred parameters can be assumed ready for activation from a minimum applicable time after the completion of model delivery/transfer

# Contact Information

Please feel free to add/update/correct the contact information if needed

|  |  |  |
| --- | --- | --- |
| Company | Name | Email |
| Moderator | Zhihua SHI | szh@oppo.com |
| Indian Institute of Technology Madras (IIT Madras) | Anil Kumar Yerrapragada  Sai Prasad Pirati | venkatasiva@5gtbiitm.in |
| CATT | Yongqiang Fei | feiyongqiang@catt.cn |
| Baicells | Xiang YUN  Xiaonan WANG | [yunxiang@baicells.com](mailto:yunxiang@baicells.com)  wangxiaonan@baicells.com |
| Lenovo | Jianfeng Wang  Vahid Pourahmadi | wangjf20@lenovo.com  vpourahmadi@lenovo.com |
| ZTE | Xingguang WEI | [wei.xingguang@zte.com.cn](mailto:wei.xingguang@zte.com.cn) |
| Xiaomi | Qin MU | muqin@xiaomi.com |
| NTT DOCOMO | Haruhi Echigo | haruhi.echigo.fw@nttdocomo.com |
| Fujitsu | Xin WANG | wangxin@fujitsu.com |
| ETRI | Yongjin Kwon | yjkwon@etri.re.kr |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |
| Panasonic | Hidetoshi Suzuki | suzuki.hidetoshi@jp.panasonic.com |
| Samsung | Ameha | amehat.abebe@samsung.com |
| Spreadtrum | Hualei Wang  Shijia Shao | [Hualei.wang@unisoc.com](mailto:Hualei.wang@unisoc.com)  Shijia.shao@unisoc.com |
| AT&T | Isfar Tariq | Isfar.tariq@att.com |
| CMCC | Yuhua Cao  Yi Zheng  Dan Song | [caoyuhua@chinamobile.com](mailto:caoyuhua@chinamobile.com)  [zhengyi@chinamobile.com](mailto:zhengyi@chinamobile.com)  songdan@chinamobile.com |
| NEC | Peng Guan  Pravjyot Singh Deogun  Jiang Yi | guan\_peng@nec.cn  pravjyot.deogun@EMEA.NEC.COM  y-jiang\_ct@nec.com |
|  |  |  |
| CEWiT | Ebin Chacko  Shiv Shankar | [echacko@cewit.org.in](mailto:echacko@cewit.org.in)  shivshankar@cewit.org.in |
| Intel | Debdeep Chatterjee | debdeep.chatterjee@intel.com |
| NVIDIA | Xingqin Lin | [xingqinl@nvidia.com](mailto:xingqinl@nvidia.com) |
| CICTCI | Rui Zhao | [zhaorui@cictci.com](mailto:zhaorui@cictci.com) |
| Continental Automotive | Hojin Kim | hojin.kim@continental-corporation.com |
| LG Electronics | Jiwon Kang | jw.kang@lge.com |
| LG Electronics | Haewook Park | haewook.park@lge.com |
| Futurewei | Chunhui (Allan) Zhu | chunhui.zhu@futurewei.com |
| MTK | Pedram  Yu-Jen  harrison | [pedram.kheirkhah@mediatek.com](mailto:pedram.kheirkhah@mediatek.com)  [yu-jen.ku@mediatek.com](mailto:yu-jen.ku@mediatek.com)  harrison.chuang@mediatek.com |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| Mavenir | Fan Yang  Yuanlong Yang | [fan.yang@mavenir.com](mailto:fan.yang@mavenir.com)  yuanlong.yang@mavenir.com |
| Ruijie | Ke Zhong | zhongke@ruijie.com.cn |
| Meta | Avik Sengupta | aviksg@meta.com |

# Appendix B: Reference/tdocs

1. R1-2500052 Discussion on other aspects of AI/ML model and data on AI/ML for NR air-interface FUTUREWEI
2. R1-2500070 Discussion on other aspects of AI/ML model and data ZTE Corporation, Sanechips
3. R1-2500152 Discussion on other aspects of the additional study for AI/ML Huawei, HiSilicon
4. R1-2500162 Discussion on other aspects of AI/ML model and data Spreadtrum, UNISOC
5. R1-2500205 Further study on AI/ML for other aspects CATT, CICTCI
6. R1-2500255 Discussion on other aspects of AI ML model and data China Telecom
7. R1-2500278 Discussion on other aspects of AI/ML model and data CMCC
8. R1-2500341 Other aspects of AI/ML model and data vivo
9. R1-2500392 Discussion on other aspects of AI/ML Ericsson
10. R1-2500469 Additional study on other aspects of AI/ML model and data OPPO
11. R1-2500549 AI/ML Model and Data Google
12. R1-2500559 Discussions on other aspects of AlML In NR air interface TCL
13. R1-2500568 Discussion on other aspects of AI/ML model and data LG Electronics
14. R1-2500591 Discussion on other aspects of AI/ML model and data NEC
15. R1-2500639 Discussion on other aspects of AI/ML model and data Lenovo
16. R1-2500690 Additional study on other aspects of AI model and data NVIDIA
17. R1-2500714 Further study on AI/ML model and data Xiaomi
18. R1-2500770 Discussion on other aspects of AI/ML models and data Apple
19. R1-2500815 Discussion on the terminology alignment TR in SA Panasonic
20. R1-2500838 Views on additional study for other aspects of AI/ML model and data Samsung
21. R1-2500904 Discussion on other aspects of AI/ML model and data ETRI
22. R1-2500929 Discussion on other aspects of AI/ML model and data Fujitsu
23. R1-2500974 Other aspects of AI/ML for two-sided model Nokia
24. R1-2500976 Discussion on other aspects of AI/ML model and data Continental Automotive
25. R1-2501079 Other Aspects of AI/ML framework AT&T
26. R1-2501148 Other aspects of AI/ML model and data Qualcomm Incorporated
27. R1-2501194 Discussion on other aspects of AI/ML model and data NTT DOCOMO, INC.