**3GPP TSG RAN WG1 #118 R1-24xxxxx**

**Maastricht, NL, August 19th – 23rd, 2024**

Source: Moderator (OPPO)

Title: Summary#1 for other aspects of AI/ML model and data

Agenda Item: 9.1.3.3

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 agenda item 9.1.3.3 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 captured in the Rel-19 WID (RP-213599) is 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 [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] | *Proposal 1: For MI-Option 1, conclude that an associated ID is valid only within a cell, and the network assigns/manages associated IDs. Associated ID for multiple cells is beyond the scope of RAN1 discussion.*  *Observation 1: For MI-Option 1, one associated ID may be mapped to multiple models trained using the data collected based on the associated ID.*  *Proposal 2: Conclude that associated ID is not model ID.*  *Proposal 3: For MI-Option 1, considering the model identification perspective, conclude that RAN1 only discuss the case that model IDs are assigned in procedure AI-Example1.*   * *The procedure without Step D is not considered in R19 as model identification and it can be a separate discussion (e.g., for alignment of additional conditions).*   *Proposal 4: For MI-Option 1, conclude that model IDs are assigned only by the NW (Alt. 1).*  *Proposal 5: For MI-Option 1, revise Step C of AI-Example1 procedure as below.*   * *C: AI/ML models are developed (e.g., trained, updated) at UE side based on the collected data corresponding to the associated ID(s) and cell ID/information.*   *Proposal 6: For MI-Option 1, further discuss whether associated ID and cell ID/information need to be sent to the UE that collects data.*  *Proposal 7: Clarify the following for MI-Option 2.*   * *The boundary between MI-Option 1 and MI-Option 2, as both options are related to data collection/dataset.* * *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 8: Study the following, if MI-Option 2 is supported.*   * *Method of referring to a dataset, e.g., whether we can use the associated ID in MI-Option 1.* * *Necessity of dataset transfer and the mechanism of doing it.*   *Proposal 9: For MI-Option 2, conclude that UE-side additional condition(s) do not need to be considered.*  *Proposal 10: For MI-Option 2, conclude that model IDs are assigned only by the NW.*  *Proposal 11: Support MI-Option 3 with further study of its procedures and specification impact, based on model transfer Case z4.*  *Proposal 12: Consider MI-Option 4 as a valid option only if the relationship between the reference model and multiple derived models can be clarified.*  *Proposal 13: Model identification option IM-Option5 is not pursued for Rel-19 normative work.* |
| Spreadtrum[2] | *Observation 1: MI-Option2/3/4 can be considered for two-sided use cases*  *Proposal 5: For AI-Example of MI-Option 1, it is up to NW to assign the model ID.*  *Proposal 6: MI-Option 5 can be deprioritized.*  *Proposal 7: For two-sided model, model identification, e.g., MI-Option2/3/4, can be considered, at least for the sake of providing pairing of two-sided models.* |
| Google[3] | *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 communication 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 3: Deprioritize MI-Option 2 and MI-Option 3.*  *Proposal 5: One associated ID can be mapped to one or multiple model ID(s)*   * *For UE-side model, the NW only configures the associated ID* * *For two-side model, the NW and UE should maintain the same understanding on the model ID for model inference*   + *Model ID can be either configured by the NW or reported by the UE*   *Proposal 6: Support the NW to configure whether the associated ID in different cells indiciate the same additional conditions or not.* |
| Tejas[4] | *Proposal 1: NW assigns the Model ID associated with the dataset and this is transferred from NW to UE (as shown in AI-Example2-1 step A).*  *Proposal 2: Regarding AI-Example1 of MI-Option1,*   * *Global Cell Identity (GCI) can be used as an associated ID.*   *Proposal 3: Regarding the relationship between model ID(s) and the associated ID(s) in AI- Example1 of MI-Option1, one associated ID(s) can be linked to multiple model IDs.*  *Proposal 4: Regarding AI-Example1 of MI-Option1*   * *Down select to Alt1 (i.e., NW assigns Model ID)* |
| CMCC[5] | *Observation 1: There are two different directions for UE-sided model(s) development in MI-Option1:*   * *Direction 1: Step A+B+C*   + *It mainly is used for the procedure of functionality based LCM* * *Direction 2: Step A+B+C+D*   + *It mainly is used for the procedure of model-ID based LCM*   *Observation 2: If only associated ID in MI-Option1 is indicated to UE-side, and UE-side does not know the detailed meaning of NW-side additional conditions, then the consistency of NW-side additional conditions can only be maintained per cell or cell group.*  *Observation 3: One model could correspond to multiple associated IDs, or one associated ID could correspond to multiple associated IDs in MI-Option1.*  *Proposal 1: The following aspects could be the starting point when discussing the information of model during model identification:*   * *The related functionality/AI enabled feature of model* * *Model’s applicable scenarios, configurations* * *Type/dimension of model input/output*   *Proposal 2: The following alternatives can be considered to resolve the NW-side additional consistency issue over multi cells or cell groups without Step D in MI-Option1:*  *• Alt 1: Model transfer/delivery when UE need to move another cell or cell group*  *• Alt 2: NW-side additional conditions indication to UE side*  *• Alt 3: Offline inter-vendor collaboration, including gNB-gNB and/or gNB-UE collaboration*  *Proposal 3: It is suggested to deprioritize Alt 3 and Alt 4 for model ID(s) determination/assignment in MI-Option1:*   * *Alt.3: Associated ID(s) is assumed as model ID(s)* * *Alt.4: Model ID is determined by pre-defined rule(s) in the specification*   *Proposal 4: It is suggested to further study Alt 1 and Alt 2 for model ID(s) determination/assignment in MI-Option1:*   * *Alt.1: NW assigns Model ID* * *Alt.2: UE assigns/reports Model ID*   *Proposal 5: Confirm the working assumption:*  *Working Assumption@ RAN1# 117*  *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)*   *Proposal 6: Regarding the associated ID for Rel-19, the UE can assume that NW-side additional conditions with the same associated ID are consistent for multiple cells if they are within the same infrastructure vendor*  *Proposal 7: There maybe one dataset ID associated with the transferred dataset in the Step A in MI-Option2.*  *Proposal 8: 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 9: 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 10: For MI-Option 3, it may include the following procedure:*   * *Step1: NW may transmit the owned or configurable model list to UE.* * *Step2: UE will report supported model list to the NW.* * *Step3: NW may transfer model and assign the model ID to UE side for the following model deployment, model inference and corresponding LCM operation.*   *Proposal 11: It is suggested to deprioritize MI-option 4 for model identification.* |
| Intel[6] | *Functionality and Model Identification*  *Observation 1:*   * *Model-ID-based identification is a necessary component to support:*   + *Model transfer from network to UE.*   + *Pairing of two-sided models.* * *Model-ID-based identification can be instrumental in enabling efficient means for alignment between network and UE to ensure consistency between training and inference.*   *Observation 2:*   * *In the context of Life Cycle Management (LCM) for AI/ML models/functionality, compared to functionality-level identification, model-level identification offers finer granularity of access and control for various LCM aspects in terms of performance expectations, performance monitoring, and subsequent decision making that affect model update, model switching, model (de-)activation, at the likely cost of increased exposure of underlying model(s) to serve a given AI/ML functionality.*   *Proposal 1:*   * *Consider support of model-ID-based identification by enabling provision of model ID to a UE by the network for model identification type B.*   + *Model-ID-based identification can apply for all three model identification options (MI-Options 1, 2, 3) subject to support of dataset transfer and model transfer for MI-Options 2 and 3 respectively.*   *Proposal 2:*   * *For MI-Option 1, on determination/assignment of model ID(s), the following options are considered further:*   + *(Opt. A) Model(s) ID(s) are already determined/assigned prior to assignment of Associated ID.*   + *(Opt. B) Model(s) ID(s) are assigned/determined at the time of association to the configuration(s) and/or indication(s), i.e., following assignment of Associated ID.*   *Observation 3:*   * *[Relationship between model ID(s) and Associated ID(s) for Alt. 1/2/4] For MI-Option 1, if model ID(s) are already assigned/determined prior to assignment of Associated ID,*   + *the assigned/determined model IDs for the models reported by the UE are associated with the assigned Associated ID(s) without any inherent relationship between model ID(s) and the Associated ID;*   + *a single model, identified by a model ID, may map to one or multiple Associated ID(s);*   + *multiple models, identified by respective model IDs, may map to one Associated ID;*   + *it is possible that none of the identified models may be reported in response to assignment of an Associated ID for a given data collection configuration/indication.*   *Observation 4:*   * *[Relationship between model ID(s) and Associated ID(s) for Alt. 1/2/4] For MI-Option 1, if model ID(s) are assigned/determined at the time of association to the configuration(s) and/or indication(s), i.e., following assignment of Associated ID,*   + *model ID(s) for the reported model(s) can be determined/assigned to have a hierarchical relationship to an Associated ID, i.e., follow Associated ID;*   + *one or multiple models, identified by respective model ID(s), may map to one Associated ID.*   *Observation 5:*   * *For MI-Option 1, if Associated ID is assumed as model ID,*   + *multiple physical models may be mapped to an Associated ID corresponding to a set of configuration(s) and/or indication(s) for data collection and share a common model ID;*   + *depending on the number of models reported for an Associated ID, this alternative may provide a level of control for LCM operations that lie on the continuum between functionality- and model-level LCM.*   *Proposal 3:*   * *MI-Option 1 can be applicable and beneficial for all the identified use-cases considered during Rel-19 (beam management, positioning, CSI prediction, and CSI compression) that would benefit from model-level granularity for LCM operations for a given functionality.*   *Proposal 4:*   * *For MI-Option 2 for UE part of two-sided model, model IDs are UE-specific.* * *For MI-Option 2 for UE part of two-sided model, model IDs can be either assigned/reported by UE or be determined based on a specified relationship with respect to dataset ID.* * *For MI-Option 2 for UE part of two-sided model, one or more model IDs may be mapped to a single dataset ID.*    + *In case of multiple model IDs, they can correspond to different models developed under different sets of assumptions on UE-side additional conditions that may be transparent to the NW.*   *Proposal 5:*   * *For MI-Option 2 for UE part of two-sided model or UE-sided model, to alleviate the reliance on inter-vendor coordination on characteristics of dataset(s) that may be transferred from NW/NW-side to a UE, candidate values for certain characteristics may be specified. Examples of such characteristics include:*   + *format for data representation in a dataset, including numbers of model inputs, outputs, quality, and associated labels (as applicable),*   + *size of a dataset,*   + *specific details on the data representation, including aspects like normalization-related information, etc.*   *Proposal 6:*   * *MI-Option 2 can be applicable and beneficial for:*   + *two-sided models for CSI compression use-case,*   + *UE-sided models for positioning use-case Case 1 for which dataset with measurements and associated ground-truth labels (location coordinates) can be transferred/delivered from LMF to UE for model training at the UE (or UE-side OTT server),*   + *localized (site-/cell-specific) models trained at UE-side (or UE-side OTT server).*   *Proposal 7:*   * *For MI-Option 3, UE-sided model or UE part of two-sided model is trained by NW and UE performs model identification procedure to request a model and its corresponding ID from NW. The model ID can further be used for model management.* * *Model transfer, along with model identification, can be provisioned to a UE by the network in response to an explicit or implicit model request from a UE. Details FFS.* * *If associated dataset for the transferred and identified model is provided by the network, then such association between dataset ID and model ID could be included as well.*   + *Alternatively, if dataset is collected at the UE side, configuration(s) and/or indication(s) for data collection could also be conveyed to the UE by the network.* * *MI-Option 3 can be applicable and beneficial for:*   + *two-sided models for CSI compression use-case,*   + *UE-sided model for which the model is trained at the network side,*   + *localized (site-/cell-specific) models trained at network side.* |
| ZTE[7] | *Observation 1: Comparison of three flavours of “MI-Option1”is as following.*   |  |  |  | | --- | --- | --- | | *Flavour* | *Detailed steps* | *Analysis* | | *Flavour#1* | *Step A + Step B + Step C + Step D1*  *Step D1: UE reports the Functionality information to the base station based on the associated ID indicated by the base station and availability of UE’s model.* | *UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.*  *Based station performs Functionality-based LCM based on the Functionality information reported by the UE* | | *Flavour#2* | *Step A + Step B + Step C + Step D2*  *Step D2: UE reports the Functionality information and associated ID supported by the UE to the base station based on the availability of UE’s model.* | *UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.*  *Based station performs Functionality-based LCM based on the Functionality information reported by the UE*  *The difference between Flavour#1 and Flavour#2 is that, in addition to the Functionality information, UE also indicates the supported associated ID to the base station with Flavour#2.*  *The advantage of Flavour#2 is that, once base station updates the associated ID, UE doesn’t need to report the supported Functionality again. Base station can determine the supported Functionality of the UE by the reported associated ID from UE.* | | *Flavour#3* | *Step A + Step B + Step C + Step D3*  *Step D3: UE reports the Functionality information and associated ID supported by the UE to the base station based on the availability of UE’s model. Model ID is assigned by the base station. Model-ID based LCM is performed.* | *UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.*  *After assigning the model-ID, based station performs model-ID based LCM.* |   *Proposal 1: If model ID is needed for MI-Option1, model ID is assigned by the base station instead of UE.*  *Observation 2: MI-Option 1 can’t address the model pairing issue for two-sided model.*  *Observation 3: The detailed procedure of associated ID is related to applicable functionality report.*  *Proposal 2: Regarding MI-Option2, dataset ID is considered as model ID.*  *Proposal 3: Regarding MI-Option2, the transmission overhead of dataset transfer needs to be addressed.*  *Observation 4: Regarding MI-Option 4,*   * *Standardization of reference UE-part model is preferred.* * *There is no such issue as multi-vendor collaboration and model pairing if reference UE-part model is standardized.*   *Observation 5: Regarding MI-Option 5, more clarification is needed.*   * *It can be applied to address the additional condition issue, but not for multi-vendor collaboration and model pairing.*   *Proposal 4: In Rel-19 AI/ML framework study, type B model identification is prioritized compared with type A model identification.*  *Proposal 5: In Rel-19 AI/ML framework study, the study of model identification should focus on the two-sided model instead of one-sided model.*  *Proposal 6: In Rel-19 AI/ML framework study, in order to support a complete and unified solution for model identification, multi-vendor collaboration, and model pairing, MI-Option 2, MI-Option 3, and MI-Option 4 are prioritized.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | *Category* | *Model identification* | *Multi-vendor collaboration* | *Whether model pairing is addressed* | *Analysis* | | *Data collection related configuration(s) and/or indication(s)* | *MI-Option 1* |  | *No* | *MI-Option 1 may not address the model pairing issue directly. For example, even with the same data collection related configurations and/or indications, the UE-part model may not be compatible with the network-part model due to the quantization method, data resolution, output size of the CSI generation part, etc.* | | *Dataset* | *MI-Option 2* | *Option 2, Option 4* | *Yes* | *MI-Option 2 (dataset transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue.* | | *Model transfer* | *MI-Option 3* | *Option 3, Option 5* | *Yes* | *MI-Option 3 (model transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue.* | | *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.* | | *Model monitoring* | *MI-Option 5* |  | *No* | *MI-Option 5 (model monitoring) may not address the model pairing issue directly. For example, in case of poor model performance, it is not clear whether it is due to the incompatibility of the model or other potential reasons, e.g., additional condition.* | |
| Continental Automotive[8] | *Proposal 1: Study specifics of configuring mapping relationship between datasets and model IDs for the captured MI-option2.*  *Proposal 2: Support Model ID as mandatory to be applied as basis for model identification related issues.* |
| Ericsson[9] | *Observation 1 Information and/or indication on NW-side additional conditions and NW configuration by the NW can be considered as implicit model identification initiated by the NW.*  *Observation 2 The applicability signalling of a functionality in a reactive approach mitigates the need for model-ID based LCM.*  *Observation 3 RAN1 should study what UE-side additional condition(s) should be specified to allow categorizing the collected data based on UE-sided condition for training two sided models based on dataset delivery from NW to UE.*  *Observation 4 For UE side part training of two-sided model based on dataset exchange from NW to UE, the over-the-air delivery method has high complexity, and the feasibility of over-the-air delivery is questionable.*  *Observation 5 In model identification option 2, example 2-1, without access to the NW decoder output, end-to-end performance verification of two-sided model is challenging, and further evaluations are needed to confirm possibility to guarantee and maintain compatibility with NW decoder without such information.*  *Observation 6 If an initial common encoder is not used when training the NW-side model of different network vendor, each UE/chipset vendor would need to train different CSI generation models for different NW vendors, and when operation in the field, a UE needs to be able to load different CSI generation models depending on at least which network vendor it is connected to.*  *Observation 7 For UE side part training of two-sided model based on dataset exchange from NW to UE, "vendor-vendor specific” conformance testing would be needed to ensure robust performance (i.e., 3GPP-level multi-vendor interoperability cannot be maintained)*  *Observation 8 For UE side part training of two-sided model based on dataset exchange from NW to UE, a reference model can be used to perform end-to-end performance validation of the actual UE-side model of the two-sided model case that may be developed/optimized using the data set delivered from the NW side.*  *Observation 9 For UE side part training of two-sided model based on dataset exchange from NW to UE, a reference model allows developing a single encoder at the UE side that is compatible with multiple NW decoders.*  *Observation 10 RAN1 has yet not concluded on the support of training collaboration type 1 and therefore, there is still uncertainty in the need to support Model identification in model transfer from NW to UE.*  *Proposal 1 Conclude that information and/or indication on NW-side additional conditions by the NW can be considered as implicit model identification initiated by the NW.*  *Proposal 2 For MI-Option 1, in step A, consider the following use-case information as the “data collection related configuration(s) and it/their associated ID(s)” which is transmitted from NW to UE,*  *a. Beam management: data collection related configuration(s) comprises the set A/B configuration, and the associated identifiers comprises consistency information of the NW transmission parameters when transmitting set A/B.*  *b. Positioning: data collection related configuration(s) and their associated IDs, including:*  *i. The validity area of the model for inference, e.g., a list of TRPs that transmits PRS*  *ii. TRP/ARP location information*  *iii. Spatial domain information of PRS*  *iv. Time synchronization information of the TRPs*  *v. PRS configuration information*  *Proposal 3 For MI-Option 1, AI-example1, conclude that step A/B/C is the baseline method. The need for step D is not justified.*  *Proposal 4 For MI-Option 1, add a note that “Associated ID(s) can be considered as a logical model ID(s)”, and how the UE maps the associated ID to a possible physical model ID is transparent to the NW.*  *Proposal 5 For MI-Option 1, further study its applicability to the two-sided use case.*  *Proposal 6 RAN1 to not consider UE side part training of two-sided model based on dataset exchange from NW to UE using over the air signaling, due to:*  * High complexity of over the air signaling of the dataset*  * 3GPP-level multi-vendor interoperability cannot be maintained and vendor-vendor specific” conformance testing would be needed to ensure robust performance.*  * UE/chipset vendor would need to train different CSI generation models for different NW vendors, and when operation in the field, a UE needs to be able to load different CSI generation models depending on at least which network vendor it is connected to.*  *Proposal 7 For UE side part training of two-sided model based on dataset exchange from NW to UE, to minimize inter-vendor collaboration and preserve interoperability, a reference model should be standardized. Using a reference encoder model, the UE-side can train and validate the actual UE-side model of the two-sided model case that may be developed/optimized based on the additional dataset and without exposing the actual other part model implemented at the NW.*  *Proposal 8 For UE side part training of two-sided model based on reference model and additional dataset exchange from NW to UE, the pairing ID composes of reference model ID and a dataset ID. The data set ID is generated locally by the NW and dependent on unique vendor/location/site IDs and additional part that is proprietary generated by the NW vendor.*  *Proposal 9 For Ml-Option 2,3, and 4, RAN1 to conclude that they are not applicable for the UE-sided model use cases.*  *Proposal 10 For Ml-Option 2,3, and 4, RAN1 to conclude that there is no need to discuss until further progress is made for the two-sided CSI compression use case.*  *Proposal 11 MI-Option 5 should not be considered further.* |
| vivo[10] | *Observation 1: Associated ID and model ID have different underlying logic.*   * *Associated ID represents certain NW-sided implementation/configurations and/or wireless channel environments.* * *Model ID represents certain AI/ML model implementation, which may require additional control/awareness of model beyond associated ID.*   *Observation 2: Directly using associated ID as model ID is not future-proof for cases where real model-level awareness is needed.*  *Observation 3: Model identification for one sided model could achieve refined model level control and might be needed in the following cases.*   * *Model switching timeline alignment across two sides;* * *Model selection with appropriate performance target and complexity tradeoff;* * *Model monitoring metric calculation.*   *Proposal 1: Study the necessity of the following scenarios if model identification is to be further studied:*   * *Model switching timeline alignment across two sides;* * *Model selection with appropriate performance target and complexity tradeoff;* * *Model monitoring metric calculation.*   *Proposal 2: For the purpose of addressing the issue of maintaining consistency between training and inference with data collection related configuration(s) and/or indication(s), associated ID can be supported as the starting point.*  *Observation 4: Global associated ID may expose deployment choices of NW side, but is useful information to maintain consistency between training and inference.*  *Observation 5: 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 3: 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 4: Associated ID + cell ID(s) can be supported to indicate the applicable cell(s) for multiple cell scenario.*  *Proposal 5: Associated ID + time stamp information can be supported to indicate the applicable period of the associated ID.*  *Observation 6: ID of transferred dataset (if feasible) is not the same as the ID for model identification based on similar reasons as above for associated ID.*  *Observation 7: 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: Reference models may not need to be identified based on explicit model identification procedure, but IDs can still be associated with specified reference models to facilitate model-level LCM.*  *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.*   *Proposal 11: How MI-Option 5 (model identification via model monitoring) works is not clear.* |
| OPPO[11] | *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:*  *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 3:*  *For model identification type B MI-Option 2,*   * *Strive for achieving the 1-to-1 mapping between dataset and 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 4:*  *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 5: Functionality ID can be used for indication functionality between NW and UE.*  *Proposal 6: At least for LCM with non-3GPP-based model transfer, Local model ID can be a simple number, 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.* |
| Xiaomi[12] | *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: The necessity of Type B MI-Option 2 is weak*  *Proposal 4: The associated ID is not equivalents to the model ID*  *Proposal 5: Support cell-group unique associated ID to balance the complexity on UE side and proprietary deployment preservation on NW side*  *Proposal 6: 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 7 : 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 8: Associated ID can be considered for data collection for type A model identification* |
| Fujitsu[13] | *MI-Option1*  *Observation-1:*   * *The working assumption in RAN1#117 can be confirmed for the cell-specific model.*   *Observation-2:*   * *The associated ID without association with real NW additional conditions has difficulties to support the development of the generalized model over various physical NW additional conditions.* * *The cell-level associated ID is hard to be applied across cells.*   *Observation-3: For the cell-level associated ID, the following issues need to be clarified:*   * *The availability of cell-level models for a UE with measurement latency requests in mobility scenarios.* * *The feasibility of local model (e.g. cell-specific-model) development for a huge number of GCI.*   *Observation-4: If only the cell-level associated ID is concluded in the MI-Option 1, model identification is not necessary.*  *Observation-5: The relationship between the model ID and the associated ID needs to be further studied for the development of the generalized model.*  *Proposal-1: MI Option1 is suggested to be further studied in Rel-19 with focus on the following issues:*   * *Wider-range-applicable associated ID to ensure the training-inference consistency of NW-sided additional conditions across cells.* * *Model identification details for the generalized/global model.*   *MI-Option2*  *Observation-6: Regarding AI-Example2-1, dataset indication information can be considered for dataset transfer with at least the following two assumptions:*   * *Alt-1: Dataset ID(s) is assumed as model ID(s).* * *Alt-2: A model can be trained with multiple datasets, and model ID of the model is associated with multiple dataset IDs which indicate the datasets for the model training.*   *Proposal-2: MI-Option2 can be further studied for the two-sided model in Rel-19 with focus on the following aspects:*   * *Dataset indication.* * *The relationship between dataset ID and model ID.* * *The details of model identification.*   *MI-Option3*  *Proposal-3: MI-Option3 can be further studied for the two-sided model in Rel-19 with focus on the following aspects:*   * *Indication of the model structure.* * *Indication of the transferred model parameters.* * *The relationship between model structure indication, model parameters indication and model indication.*   *MI-Option5*  *Proposal-4: For MI-Option5, the model ID assigned from NW is for identifying the model’s applicability under certain NW-side additional conditions.*   * *It can be named as model applicable ID.* * *It can be assumed as local ID.* * *It is applicable to the model performance assessment after model transfer.*   *Proposal-5: The procedures of MI-Option5 are further clarified as below:*   * *If model applicable ID(s) of a cell is unavailable, the UE initiates the model monitoring/selection procedure, and the NW provides necessary measurement configurations accordingly.*   + *Applicable model(s) is selected via model monitoring under certain NW-side additional conditions.*   + *The NW assigns model applicable ID(s) to the selected model(s).* * *If model applicable ID(s) of a cell is available, the UE reports the model applicable ID(s) of a cell to the NW. The NW can decide activation of the corresponding model and skip the model monitoring/selection procedure.*   *Proposal-6: MI-Option5 is suggested to be further studied in Rel-19 with focus on:*   * *Performance monitoring/assessment schemes, including the specific schemes for the post-deployment performance monitoring and the post-model-transfer performance monitoring.* * *Applicable ID details.*   *Summary of the considerations on model identification*  *Observations-7: Model identification is necessary for the two-sided model when taking the following issues into account:*   * *Dataset and its related model part indication in MI Option2.* * *Model/model structure identification in MI Option3.* * *Model pairing request (TR38.843 of Rel-18 SI [2]).*   *Observations-8: Model identification can be used for the one-sided model to address the following issues:*   * *To indicate a generalized model across cells.* * *To indicate the known model structure in model transfer of a UE-sided model.* * *To indicate the applicability of a local model/model part through model transfer.*   *Proposal-7: Model identification is necessary for the two-sided model. If CSI compression is justified to be moved into the normative phase, model identification is suggested to be considered in its normative work.*  *Proposal-8: Model identification and the associated ID beyond cell-level are important to address the consistency issue across cells. Further study on the details of model identification and the wider-range-applicable associated ID can be considered in BM and PO sub agendas.* |
| CATT[14] | *Observation 1: Generalization capability and performance monitoring are considerable alternative solutions to address/alleviate additional condition consistency issue and provide minimum guaranteed performance.*  *Observation 2: Unless there is a dataset identifier (e.g. dataset ID) acknowledged across different cells, the dataset can only be assumed corresponding to the originated cell that transfers the dataset, so as the model ID(s). This implies localized (cell-specific) model(s) have to be assume, which may not be reasonable.*  *Observation 3: 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 Rel-18 AI/ML-based CSI compression.*  *Proposal 1: Offline model identification, i.e. type A, is out of 3GPP and cannot be justified by RAN1.*  *Proposal 2: Unless clear additional benefit is found and justified for MI-Option1 compared to functionality-based LCM with associated ID, do not support MI-Option1.*  *Proposal 3: In AI-Example1, model ID is assumed to be independent from associated ID.*  *Proposal 4: Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent within:*   * *One cell (baseline);* * *One cell group. Whether/how to categorize cells into a cell group is up to NW implementation;* * *Other ranges (e.g. W vendor, per PLMN or global) are not recommended.*   *Proposal 5: In AI-Example1, model ID is assigned by network after UE reporting the information of its AI/ML models to the network.*  *Proposal 6: In AI-Example1, meta information (if supported) of an AI/ML model carries all related associated ID(s) of the AI/ML model, and is transmitted from UE to network.*   * *Meta information may carry other information, which is up to future discussion.* * *The procedure and signaling of meta information transmission is out of RAN1.*   *Proposal 7: MI-Option2 is only discussed under the context of two-sided model use case.*  *Proposal 8: In MI-Option2, study dataset ID and its applicable range to clarify whether dataset can be uniquely identified across different cells.*  *Proposal 9: In AI-Example2-1, as a starting point, the mapping relationship between dataset 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 dataset and model ID is one-one-one mapping, and how to determine model ID in this case.*   *Proposal 10: In AI-Example2-1, for AI/ML-based CSI compression, the need and benefit of UE-side additional condition(s) requires further justification.* |
| TCL[15] | *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.*  *Proposal 1: 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.* |
| LG[16] | *Proposal#1. Clarify that any LCM that does not require assigning model ID belongs to functionality-based LCM.*  *Proposal#2. In AI-Example1, Step A/B/C are sufficient for UE-sided model to address the consistency issue between training and inference.*  *Proposal#3. For AI-Example1, Step D may be useful for two-sided model for model pairing but not for UE-sided model. In addition, the Alt3 of Step D in the examples needs further clarification on how to ensure one-to-one mapping between associated ID and each UE-sided model.*  *Proposal#4. On the necessity of model identification and model-based LCM, conclude that*   * *they are necessary for model transfer (if supported) and two-sided model cases (if supported).* * *they are not necessary for one-sided model cases.*    + *for one-sided model cases, other means to provide information/indication for scenario/site-specific models can be considered under functionality-based LCM framework.* |
| Lenovo[17] | *Proposal 1: Confirm the Working Assumption that the UE assumes that NW-side additional conditions with the same associated ID for the data collection are consistent at least within a cell.*  *Proposal 2: Data collection configuration(s) can be composed of the set of conditions/additional conditions of the UE, of the gNB, and even of other nodes in the network affecting the measured data, where each conditions/additional conditions can be represented using a separate associated IDs, or a single ID associated to a particular combination of conditions/additional conditions.*  *Proposal 3: Study the association between the data collection related configuration(s) with an associated ID and the dataset.*  *Proposal 4: Associate an ID with the dataset to be transferred, and further exchange the information on the models developed by the dataset based on the associated ID. The relation between dataset associated IDs and data collection related configuration(s)/associated ID can be further studied.*  *Proposal 5: Further study MI-Option 3 with inter-vendor training collaboration Option 3 (Standardized reference model structure + Parameter exchange between NW-side and UE-side) and Option 5 (Standardized model format + Reference model exchange between NW-side and UE-side) in the AI/ML-based CSI compression use case.*  *Proposal 6: Further study the mechanisms to indicate the candidate model structures and the signaling on the corresponding parameters.*  *Proposal 7: The model identification procedure dedicated to IM-Option5 is not pursued for Rel-19 normative work.*  *Proposal 8: The UE sided model is not necessary to be identified, even though some information related with the AI/ML models or functionalities need to be shared between the NW and the UE.*  *Proposal 9: Further study the necessity and detailed procedure of model identification for the UE part of a two-sided model in the AI/ML-based CSI compression use case.* |
| IIT Kanpur[18] | *Proposal 2: Prioritize study of ALT-4 for Step D in cases of MI-option1 for Rel-19.*  *Proposal 3: The model identification procedure dedicated to MI-Option5 is not pursued for Rel-19 normative work* |
| NVIDIA[19] | *Proposal 1: Conclude that there is a need for model identification in the context of LCM.*  *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* |
| InterDigital[20] | *Observation 1: In beam management use case, for the cases of AIML models only at network side, the LCM procedures can be network implementation specific, and the model identification may not be necessary.*  *Observation 2: In positioning use case 3a, if LCM is performed by the LMF, NW-side model identification may be necessary since the LMF may need to know information about AIML model(s) implemented at the gNB.*  *Observation 3: Functionality-based LCM and model-ID-based LCM may be applicable for potentially different use cases, model deployments, model management granularity and collaboration levels.*  *Observation 4: Model-ID based signaling in a Functionality is beneficial for model-level management (e.g., pairing of models) of UE-part of two-sided models.*  *Observation 5: based on how to assign Associated ID, associated ID level LCM may be equivalent to model-ID based LCM. Therefore, Alt.3 seems to be enough to cover most of use case for the single-sided model developed and used at the UE side.*  *Proposal 1: An associated ID corresponds to data collection related configuration(s) and NW-sided additional conditions.*  *Proposal 2: For single-sided model developed and used at the UE side, down-select Alt-3 only for MI-Option 1 for further study.*  *Proposal 3: Associated ID signaled by the NW can be used at the UE-side to determine consistency between training and inference without additional information/configurations.*  *Proposal 4: For AIML positioning purpose, support MI-Option 1 for model identification type B.*  *Proposal 5: Clarify if MI-Option2 refers to Option 4 in the CSI compression sub-agenda item (“Option 4: Standardized data / dataset format + Dataset exchange between NW-side and UE-side”).*  *Proposal 6: Defer discussions on MI-Option2 for the UE-part of a two-sided model until more progress is made in the inter-vendor training collaboration discussion, at least for Option 4.* |
| NEC[21] | *Observation 1: Model ID is essential for use cases with model transfer, model update, or two-sided models, and is beneficial to differentiate additional conditions to ensure the consistency between training and inference.*  *Proposal 1: Support model ID and model identification in Rel-19.*  *Proposal 2: RAN1 should study following options for model identification Type B for further discussion.*  *− MI-Option 1: Model identification with data collection related configuration(s) and/or indication(s)*  *− MI-Option 3: Model identification in model transfer from NW to UE*  *Proposal 3: It is necessary to clarify the definition of associated ID for each use case separately.*  *Proposal 4: For MI-Option 1, further clarify the data collection related configurations for each use case. And support UE to request the needed data collection related configurations.*  *Proposal 5: One or more associated ID(s) can be attached to one same model ID to reflect different NW side additional conditions.*  *Proposal 6: In the model identification procedure, support the combination of Alt.1: NW assigns Model ID and Alt.2: UE assigns/reports Model ID, which is used to link a UE reported global model ID to a NW assigned local model ID.*  *Proposal 7: 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 8: 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 9: Study the grouping of cells that can ensure the consistency within a subset of cells*  *Proposal 10: Information of model monitoring methods can be provided to NW or UE. If model failure occurs, the cause of model failure may also be reported.*  *Proposal 11: Specify monitoring of inactive model/functionality for the purpose of activation/selection/switching of UE-side models/UE-part of two-sided models /functionalities for Rel-19 AI/ML.*  *Proposal 12: Discuss whether a UE can perform inference of two models/functionalities concurrently where one model/functionality is inactive but being monitored and other model/functionality is activated at UE.*  *Proposal 13: Support adaptive model/functionality selection, activation, deactivation, switching, and fallback.*  *Proposal 14: Support event triggered AI/ML functionality/model activation/deactivation/switching.* |
| Nokia[22] | *Proposal 1: For MI-Option 1, considering steps A -to - D, the following aspects are further applicable,*   * *For associated ID,*    + *Defining of associated IDs is up to the NW vendor implementations and shall not disclose any proprietary NW information.*   + *For BM use-cases, associated ID can be linked to CSI resource configuration (CSI-resourceConfig), or resource sets defined by a CSI-resourceConfig.*   + *For positioning use-cases, associated ID can be linked to the PRS resource configuration (NR-DL-PRS-Info) or PRS resource sets (nr-DL-PRS-ResourceSet) defined by a NR-DL-PRS-Info.*   + *In addition to the associated IDs, data collection configuration(s) may also associate with the global cell identities (CGIs).*   + *A fixed bit field, e.g., 8/10 bits (provides max 256/1024 IDs) can be considered for the associated ID.* * *For model-ID,*    + *In Step D, the UE assigns model ID, and reports associated IDs (and optionally CGIs) related to the assigned model ID.*      - *Model ID may be related to one or more associated IDs.*   + *Reporting of model-IDs does not have to be in the UE-capability report. RAN1/2 to investigate further exact reporting details.*   *Proposal 2: For two-sided CSI compression models, whether to consider dataset transfer or not may be discussed in AI 9.1.3.2, and RAN1 may wait for any requirements for dataset transfer before binding it with model identification.* |
| Samsung[23] | *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.*   *Proposal#3: For AI-Example2-1 of MI-Option2, support the indication associated to the 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.*   *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 functionality-based LCM, use UE capability report for the UE to indicate to the gNB, the number of models it runs for all functionalities within a boundary condition. All functionalities activated within one boundary condition are counted as 1 or , where is reported by the UE.*  *Proposal#7: For model-ID based LCM, UE may report the maximum number of simultaneously active models it supports.*  *Proposal#8: 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#9: 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.* |
| Panasonic[24] | *Observation 1: Model development is not a UE but UE side. Model compilation and test are more realistic in UE side and not within a UE.*  *Observation 2: Availability of the model is always by a UE and not by UE side.*  *Observation 3: For dataset or model parameters available at NW, the most efficient method is from NW to UE side and not involving a UE.*  *Observation 4: For the delivery of data, dataset or model parameters available at a UE to a UE side, although U-plane over the top exchange would be available, involvement of NW has some merit/demerit to be discussed in upper layer expertise. From RAN1 perspective, there is no difference between U-plane or involvement of NW.*  *Observation 5: For the delivery of a developed/compiled/tested model at UE side to a UE, although U-plane over the top exchange would be available, involvement of NW has some merit/demerit to be discussed in upper layer expertise. From RAN1 perspective, there is no difference between Uplane or involvement of NW.*  *Proposal 1: RAN1 should not limit the option related to other than RAN1 aspect. For non-RAN1 related part, RAN1 just identifies the options and informs to upper layer WGs like RAN2 and SA2.*  *Proposal 2: For MI-Option 1 of model identification with data collection, the required interaction between UE and UE side should be described.*  *Proposal 3: For MI-Option 2 of model identification with data set transfer, the required interaction between UE and UE side should be described.*  *Proposal 4: For MI-Option 3 of model identification in model transfer, the option that model parameters are directly delivered to UE side and compiled/tested at UE side should be described. In addition, required interaction between UE and UE side should be described.*  *Proposal 6: The consistency of the associated ID is at least within a PLMN.* |
| ETRI[25] | *Observation 1: The Associated ID can be used to configure and manage datasets generated through the data collection process.*  *Observation 2: By utilizing the dataset ID, the NW can manage the dataset transfer and related information.*  *Proposal 1: Model-ID-based LCM can be integrated with Functionality-based LCM by using model ID for LCM operations.*  *Proposal 2: Associated IDs are configured and managed for each functionality.*  *Proposal 3: Associated ID and Model ID can have a many-to-many relationship.*  *Proposal 4: From RAN1 perspective, for UE-sided model(s) developed at UE side, the following procedure is an example (noted as AI-Example2-2) of MI-Option2 for further study (including the feasibility/necessity).*  *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 solutions for dataset exchange is out of RAN1 scope.*  *B: UE-sided model(s) is(are) developed based on at least the above dataset.*  *C: UE reports information about its UE-sided model(s) corresponding to the above dataset to the NW.*  *FFS: How the model ID is determined/assigned for each AI/ML model (including relationship between dataset and model ID)* |
| Apple[26] | *Proposal 1: In MI-option 1, the associated ID(s) is assumed as model ID(s).*  *Proposal 2: In MI-option 2, the dataset ID(s) is assumed as model ID(s).*  *Proposal 3: In MI-option 3, the NW assign the model ID and transferred together with model parameters.*  *Proposal 4: Association ID can be applicable for multiple cells. A list of global cell IDs can be signaled per association ID.* |
| AT&T[27] | *Proposal 1: Study the common framework for the following aspects of the different AI/ML use case.*   * *Model identification* * *Model delivery/transfer* * *Signaling for Model ID based LCM* * *Performance monitoring* * *Data collection* * *Reporting of additional conditions*   *Proposal 2: 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 3: For both model identification Type B1 and B2*   * *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:*   * *RAN1 focus on the study of MI-Option 1 (including the necessity) for the use cases of AI-based positioning, AI-based beam management.* * *RAN1 focus on the study of MI-Option 1 (including the necessity) for the use case of CSI prediction.* * *RAN1 focus on the study of MI-Option 2/3/4 (including the necessity) for the use case of CSI compression.*   *Proposal 5: 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* |
| Meta[28] | 1. *Support model ID based model identification Type B with at least MI-Options 1, 2* 2. *For MI-Option-1 support steps A-C for ensuring consistency of network side additional conditions and use Alt-3 for associating model IDs to associated ID i.e., associated ID is assumed to be model ID* 3. *For MI-Option-2, a dataset ID similar to associated ID can be used as a model ID for the UE side model or UE part of the 2-sided model and only step A-B are used in this case.* |
| DOCOMO[29] | *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: Depending on the model ID assignment in step A or C in AI-Example 2-1, model ID represents NW part model or UE part model. Preferable model ID granularity depends on how much NW manages UE side models.*  *・Model ID representing NW part model is sufficient, if NW cares about only the compatibility between UE part model and NW part model*  *・Model ID representing UE part model is preferable, if NW prefers to be aware of multiple UE part models applicable to one dataset.*  *Observation 4: 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.* |
| Sharp[30] | 1. *Associated ID is not equivalent to model ID.* 2. *Confirm the following 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)*   1. *Use both CGI and the associated ID to ensure the consistency between model training and inference across multiple cells.* 2. *Model ID should be supported at least for model ID based LCM.* |
| Huawei[31] | *Observation 1: The boundary between model identification and functionality identification for the Functionality with model ID is not clear.*  *Observation 2: The feasibility of MI-Option 1 for one-sided model is unclear, since introducing globally unique model ID for the purpose of data categorization indication is not really helpful to the UE side and may harm the proprietary preservation of the NW side.*  *Observation 3: If MI-Option 4 and MI-Option 5 need to be classified to model identification, the definition of model identification may need to be revisited.*  *Proposal 1: Consider functionality-based identification/LCM with model ID as the same category with model-ID-based identification/LCM until further clarification on the difference is achieved.*  *Proposal 2: For studying the applicable sub use cases of model identification and model-ID-based LCM, take two-sided model as the starting point.*  *Proposal 3: MI-Option 1 for one-sided model with globally unique model ID should be deprioritized.*  *Proposal 4: For ensuring consistency between training and inference via data collection related configuration(s) and/or indication(s), consider associated ID under functionality identification subject to cell specific manner.*   * *Whether to introduce the associated ID should be separately discussed under per use case.*   *Proposal 5: MI-Option 2 is applicable to two-sided model case.*  *Proposal 6: 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:*   * *Input and output of the NW side CSI generation part for training the UE side CSI generation part.* * *Other meta information, including at least: dataset ID, size of dataset, type/format of data samples, model scalability information, quantization method for CSI feedback.*   *Proposal 7: For the procedure of MI-Option 2, the model identification is achieved when the dataset ID is delivered in together with the delivered dataset.*  *Proposal 8: MI-Option 3 is applicable to two-sided model case.*  *Proposal 9: 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 10: For the procedure of MI-Option 3, the model identification is achieved when the model ID is delivered in together with the delivered model.* |
| Qualcomm[32] | *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.*  *Proposal 6: Conclude that ensuring consistency of NW-side additional conditions across different cells (for the same associated ID), is necessary to enable and facilitate the operation of UE-side AI/ML models.*   * *Further details on how to enable consistency of NW-side additional conditions across different cells should be discussed within each use case.* |
|  |  |

#### **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. |

## 1st round discussion

#### **Proposal 2.1.1**

Given companies’ inputs, we see no strong concern over the following working assumption related to associated ID. At least the associated ID can make sure the consistency in a cell-specific manner. However, we also observed that different views on whether the applicability of the associated ID across multiple cells still hold. Hence, we suggest to confirm the working assumption as it is.

**Proposal 2.1.1:**

**Confirm the 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)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | Support |
| vivo | Fine to confirm but this may need to be discussed together with Proposal 2.1.2. |
| TCL | Agree |
| Panasonic | Support |
| Spreadtrum | Support |
| CATT, CICTCI | OK |
| ZTE | We suggest to confirm the working assumption without the sub-bullet.  Firstly, this is the last the meeting of the study phase of the framework before the checking point. It is not clear what the FFS means if there is no TU left for this agenda.  Secondly, whether additional condition indication across different cells is needed should be discussed per each use case. Without the concrete understanding of what the additional condition is, the discussion in this framework agenda won’t fly. For example, if the additional condition is the so-called beam sharp, different base station have different location, heights, and other configurations, it is almost equal to cell-specific beam shape. For this case, making the consistency across cells is not practical.  **Proposal 2.1.1-ZTE:**  **Confirm the Working Assumption with the following update**  **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)~~** |
| Samsung | Ok. |
| Qualcomm | Support |
| Xiaomi | It seems the key controversial part is the FFS part, then we consider it is better to be discussed with proposal 2.1.2 together. |
| Ericsson | Support, prefer update from ZTE |
| CMCC | Support |
| Fujitsu | It can be confirmed. While the FFS part is quite important, in this sense, discussing it with proposal 2.1.2 together would be a better way. |
| NEC | Ok |
| Nokia | OK |
| LG | OK |
| Futurewei | OK. Also support the removal of the FFS. |

#### **Proposal 2.1.2**

As said before, there are many companies (including operator, UE vendor, NW vendor) believing that it is beneficial to use a single associated ID for multiple cells.

* It is a heavy burden for UE to manage the associated ID for each cell
* It is very likely there is no sufficient training data for AI model training since it will require a larger number of training data for each cell

Thus, the following proposal is suggested for further discussion:

**Proposal 2.1.2:**

**Regarding the associated ID for Rel-19, the UE can assume that NW-side additional conditions with the same associated ID are consistent within a cell group consisting of N (N>=1) cell(s) (in additional to within a cell)**

* **Note: Whether/how to categorize cells into a cell group is up to NW implementation**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | Support the proposal. |
| Huawei, HiSilicon | Do NOT support the proposal.  1) The motivation of introducing cell grouping is “It is a heavy burden for UE to manage the associated ID for each cell”, but it should not be ignored that, if cell group based associated ID is to be introduced, it would become a heavy burden of NW to manage the associated IDs across cells, which is quite challenging due to 2).  2) Considering massive factors impacting the performance (deployment scenario, down tilt angle, beam shape, beam order, codebook, etc.), it is not likely even for NW side to identify whether any two cells are the same or not. Moreover, considering it does not know the generalization capability of the UE side model, this categorization is even more difficult, assuming there is no offline co-engineering.  3) UE can perform its own data categorization methods based on data distributions across cells.  4) It is incorrect to say “**Whether/how to categorize cells into a cell group is up to NW implementation**” – UE has to know the cell grouping information, so this information has to be aligned with/informed to UE, possibly NOT by gNB but by a higher-level NW entity, e.g., OAM, CN, LMF, etc. The question is, how would other higher-level NW entities know the categorization rule of gNBs? Possibly it has to rely on the interaction with gNBs, which means it is not implementation.  Based on the analysis above, we think associated ID should be configured with cell specific manner. |
| vivo | Seems there will be strong fight on this issue. We can confirm the necessity first and then discuss mechanisms.  **Regarding the associated ID for Rel-19, mechanisms need to be designed for UE ~~can~~ to assume that NW-side additional conditions across multiple cells are consistent. ~~with the same associated ID are consistent within a cell group consisting of N (N>=1) cell(s) (in additional to within a cell)~~**   * **~~Note: Whether/how to categorize cells into a cell group is up to NW implementation~~** |
| TCL | The discussion on this proposal depends on how to encode the associate ID. If the associated ID is decided by the per-cell NW sided conditions, then it should be applicable within a cell. Otherwise, it should be applicable across multiple cells. |
| Panasonic | We support above. As whether/how to categorize cells into a cell group is up to NW implementation, it would not be possible the cell group is over multiple of PLMNs. Therefore, we propose to clarify that a cell group is only within a PLMN. |
| Spreadtrum | Support. We think that the same associated ID are consistent within a cell group is one good balance between NW vendors and UE vendors. |
| CATT, CICTCI | We are open discuss the need of one associated ID across multiple cells, but this should be a fair comparison – not only benefits for UE, but also burden for NW.  We think the note ‘Whether/how to categorize cells into a cell group is up to NW implementation’ has to be guaranteed. At the same time, HW’s view 4) displays an interesting assumption that a high-level entity should be involved to communicate with gNBs (e.g. OAM). This may be true, depends on the understanding of NW implementation: does it has to be implementation within one gNB. |
| ZTE | We don’t think this proposal is needed.  Firstly, the WID tasks this agenda to address “Necessity and details of model Identification concept and procedure in the context of LCM” instead of spending much time on additional condition.  Secondly, as mentioned in the previous proposal, whether additional condition indication across different cells is needed should be discussed per each use case. Without the concrete understanding of what the additional condition is, the discussion in this framework agenda won’t fly. Besides, this proposal will impact the ongoing discussion on the AI beam and AI pos. For example, if additional condition across cells is agreed, it will impact the discussion on whether to include additional condition indication within the CSI configuration framework. |
| Samsung | We sympathize with Huawei’s comment. While acknowledging that having consistent assumption across multiple cells on associated ID relives the UE-side burden, its feasibility and impact on performance is not clear. As a compromise RAN1 can make a conclusion that consistent assumption on associated ID across multiple cells may alleviate burden associated with data collection and model LCM on UE side  **Regarding the associated ID for Rel-19, the UE-side burden on data collection may be reduced if the UE ~~can~~ assumes that NW-side additional conditions with the same associated ID are consistent within a cell group consisting of N (N>=1) cell(s) (in additional to within a cell)**   * **~~Note: Whether/how to categorize cells into a cell group is up to NW implementation~~**   **The feasibility and performance impact is for further study.** |
| Qualcomm | Support |
| Xiaomi | * We think the necessity of consistency of NW additional condition across multiple cells should be confirmed. In our view, it is a good compromise and balance between the NW side and UE side * Regarding the potential burden or constraint mentioned by NW side, we think it could be solved by careful mapping between the network additional condition and associated ID. For example, if massive network additional factors are mapped to one associated ID, then it may be difficult to keep all the factors the same across multiple cells. But on the other hand ,if a small number of network additional factors are mapped to one associated ID, then it would be more easy to keep the consistency. |
| Ericsson | Same view as HW. Due to the high complexity in managing the IDs over multiple cells, a clear benefit needs to be concluded, which is only possible from a use case perspective.  **Updated Proposal: Regarding the associated ID for Rel-19, the need for mechanisms for UEs to assume that NW-side additional conditions across multiple cells are consistent should be handled from a use-case perspective.** |
| CMCC | Support this proposal. Putting associated ID within a cell group is a compromise between NW and UE side. And it is noted that a cell group also can only include one cell. |
| Fujitsu | We think study on beyond-cell level consistency over NW additional conditions is very important. If only the development of cell-level model is supported, how to ensure the availability of the cell-specific-models with latency restriction would be another problem in practical usage of AI/ML models besides the challenges in its development. |
| NEC | Support in principle |
| Nokia | We are not fully sure about feasibility of the proposal.  In general, NW-additional conditions are representing NW proprietary information, that is one of the reasons to implicitly indicate information. As this proposal may require coordinating (among vendors) to assign these associated ID, it may require revealing of proprietary information. |
| LG | Whether the ID can be defined per cell or per cell group would be dependent on its usage. Thus, we propose to discuss this issue per each usage of the ID. In addition, the parameter name of the associated ID can also be different per its usage, e.g. beam ID for beam use case and other ID for positioning use case, etc. |
| Futurewei | Not support at the moment; we need to first figure out what it implies. When an associated ID works beyond a cell, the management of the associated ID will go beyond a gNB. HW has listed many issues with this. Suggest we study the feasibility first before voting for the proposal. |

#### **Proposal 2.1.3**

One discussion point for AI-Example1 is the relationship between model ID(s) and the associated ID(s). The tdocs propose different options for this relationship. There are different assumptions/motivations for model identification under different options. Thus, it would be beneficial to make each option clear and then the proponent can clearly clarify necessity/benefit of model identification for their favorite option(s).

Thus, the following proposal is trying to capture these options to facilitate further discussion.

**Proposal 2.1.3:**

**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**
* **Note: Proponents of each option are encouraged to provide detailed analysis on applicable use cases, benefit, necessity and so on**

Examples of these options are shown by the following figures to facilitate the understanding of each option



Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | Support ID-Rel-Option4. One model can be trained based on the datasets collected from multiple associated IDs. Also, multiple models, which have different complexity, can be trained under the same associated ID. Hence, ID-Rel-Option4 should be supported. |
| Huawei, HiSilicon | Firstly, we have not confirmed the necessity of **AI-Example1 of MI-Option1**, since it has a substitute of Step A/B/C and additional interaction of associated IDs between UE and NW.  Secondly, the mapping between associated ID and model ID depends on the applicable area of associated ID. Assuming associated ID is cell specific, it is more likely to map multiple associated IDs to one model ID rather than the other way around.  Thirdly, the mapping relationship between the two types of IDs seems implementation, and the 4 options have no different spec impacts.  Therefore, we suggest this proposal is deferred. |
| vivo | We may need to confirm necessity of MI-Option1 first. |
| TCL | We support option 2 and 4. |
| Panasonic | Following is same comment as Fukuoka meeting.  We agree to list four options.  Our view is all option are valid and its usage depends on what level of the model needs to be identified by NW.  In ID-Rel-Option1, model ID is logical model. NW is not required to identify what physical model is actually used by UE in this case. The physical model usage is up to UE side.  In ID-Rel-Option2, model ID is physical model. For testing (or conformance test) and evaluation purpose, NW may be required to identify the physical model. The physical model can be used for multiple associated ID. This option is used for such situation.  In ID-Rel-Option3, model ID is physical model. For testing (or conformance test) and evaluation purpose, NW may be required to identify the physical model. Because of the version up of the same model or UE choose different model depending on UE judgement of the situation, one associated ID can be mapped to multiple of physical model. This option is used for such situation.  In ID-Rel-Option3, model ID is physical model. This is the case of both ID-Rel-Option2 and ID-Rel-Option3 are used. For physical model case identification case, this is more generic. |
| Spreadtrum | Support ID-Rel-Option2.  In our understanding, here model means logical model. Depending on model training methods, one model can be trained with dataset achieved by basing on one specific dataset collection configuration/associated ID, or multiple datasets achieved by basing on multiple specific dataset collection configuration/associated ID. One logical model can be mapped into one or many physical models, and it is up to implementation. |
| CATT, CICTCI | For the question itself: we think FL already lists all possible combinations. We think a model can be trained with generalization capability. At the same time, even for the same dataset, different models with different performance/complexity can be trained. Thus ID-Rel-Option4 is aligned with our understanding.  For the necessity: Same view as HW and vivo. |
| ZTE | Thanks FL for the summary. However, the following note under the previous relevant agreements should be discussed first. Since if only step A/B/C is needed, this proposal regarding step D is not needed. Thus, we suggest to first discuss whether step D is needed.  *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*  Meanwhile, one major issue left for this MI-Option1 is its usage scenarios, e.g., whether it is applicable to UE sided model and/or two-sided model. Based on our understanding, it is only applicable to UE-sided model. |
| Samsung | Ok to study the necessity of the options. It seems these options are applicable to MI-Option-2 either. |
| Qualcomm | ID-Rel-Option1 and ID-Rel-Option3 are representations of NW-initiated and UE-initiated model identification, respectively. It would be helpful if proponents of • ID-Rel-Option2 and ID-Rel-Option4 elaborate on how those options map to model identification options. The motivation is not clear to us. |
| Xiaomi | We are OK to study the necessity of the options. Currently, we consider ID-Rel-Option4 is OK for us |
| Ericsson | Share the view that necessity of MI-Option1 should first be concluded. |
| CMCC | All the options are valid at least from procedure. And ID-Rel-Option4 exactly is a superset of all the other options. It is better to first discuss the procedure of MI-Option1 and the meaning of model ID in it, then we can know which relationship is more aligned with MI-Option1. |
| Fujitsu | All options listed by FL is possible. The down-selection of the options may rely on the assumptions of associated ID in Proposal 2.1.1 and Proposal 2.1.2. |
| NEC | Ok, in principle |
| Nokia | We are not sure why to list down four variants.  It is either one-to-one mapping (model ID maps with one associated ID) or one-to-many mapping (model ID maps with more than one associated ID). All other cases do not seem to be needing any discussion when it comes to signalling. |
| LG | Fine to study, but somehow we also feel that this mapping could be different for different usage of the associated ID. It may be better if we directly discuss on a core question, i.e. whether/when we need to specify ‘model ID’ in addition to ‘associated ID’. |
| Futurewei | Agree with many companies that we are not ready for this proposal. We just want to remind the group that MI-Option1 will be valid only with ID-Rel-Option1. Otherwise, it should not be considered a method for model identification; other options of associated ID exchanges will not help identifying a model, they are just used for the alignment of additional conditions. |

#### **Proposal 2.1.4**

Another discussion point of Step D in AI-Example1 is which side is in charge of determining/assigning the model ID (if supported). Apparently, there are at least four alternatives to be further discussed and down selected. It seems time to decide how to produce the model ID. So we suggest the following proposal for discussion and potentially for down selection.

**Proposal 2.1.4:**

**For AI-Example1 of MI-Option1, study and down-select the following alternatives on determining/assigning model ID (if supported).**

* **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**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | If model ID and associated ID are the same or used for the same purpose, model ID does not need to be introduced. Then, it is reasonable to make model ID designed to enable something which associated ID cannot achieve. For Alt.1 and Alt.3, there is no benefit of introducing model ID in addition to associated D. Hence, we suggest focusing only Alt.2 and Alt.4. |
| Huawei, HiSilicon | Similar comment as Proposal 2.1.3, that we have not confirmed the necessity of **AI-Example1 of MI-Option1**. If we really need to make some progress, we suggest the description of deprioritize some alternative. Since the model ID is globally unique, it is not likely to let UE assign it or predefine it, so Alt.2 and Alt.4 can be deprioritized.  **For the study of AI-Example1 of MI-Option1, ~~study and down-select the following alternatives on determining/assigning model ID (if supported).~~ Alt.2 and Alt.4 are deprioritized.**   * **Note: the necessity of AI-Example1 of MI-Option1 has not been confirmed yet.** |
| TCL | We support option 1 and 2. |
| Panasonic | Alt1 or atl3 is baseline. Then it means only logical model. When network requires to identify the physical model, version number reported by UE as alt 2 can be additionally added. |
| Spreadtrum | Support Alt1. Like other parameters, it is better to let NW to manage model ID. |
| CATT, CICTCI | This issue may be discussed once 2.1.3 is clearer. |
| ZTE | For this proposal, we think model ID is not needed.  Firstly, there are no clear benefits to apply model ID based LCM for UE-sided model. On other contrary, due to the massive number of models from different UEs from vendors, it is not feasible for the network to manage such a huge number of model IDs.  Secondly, Rel-18, we have discussed the same issue without any outcome. The only two usage scenarios of model-ID based LCM are model transfer and model pairing. For UE-sided model, Functionality based LCM + associated ID can already address all the issues. |
| Samsung | Same view as NTT DOCOCMO. |
| Qualcomm | If down-selection is the main purpose of the proposal, we support Alt. 3. |
| Xiaomi | Can be discussed when proposal 2.1.3 is clear |
| CMCC | We prefer Alt 1. |
| Fujitsu | Considering the unclear points on the associated ID, there is no need to do down-selection now. But if the associated ID is only assumed as a cell-level ID, Alt-1 and Alt-3 are supported. |
| NEC | Support the direction of discussion. And we believe both Alt1 and Alt2 can be supported. |
| Nokia | Would be good to align on one solution than listing alternatives. Alt.3 seems to be limited way of defining the model ID. |
| LG | Agree with DOCOMO except that Alt1 may be valid if ‘model ID’ can achieve something more than the associated ID. In Alt3, we don’t need to specify ‘model ID’ in addition to ‘associated ID’. Thus, prefer to delete Alt3. |
| Futurewei | Support Alt 1 if we have to vote now.  However, we have not even decided whether model ID is necessary. We should first identify the scenarios/use cases where model IDs are needed, then determine which entity gets to assign model IDs. |

#### **Proposal 2.1.5**

**Proposal 2.1.5:**

**Conclusion:**

**From RAN1 perspective, model identification is needed and beneficial at least for the following case**

* **Two-sided model (if supported)**
  + **i.e., CSI compression (if supported)**

**Note: whether two-sided model is supported or not is separate discussion.**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | Support. |
| Huawei, HiSilicon | Support |
| vivo | Support, but not sure whether it is supported for all options of inter-vendor training collaboration.  **Conclusion:**  **From RAN1 perspective, model identification is needed and beneficial at least for the following case**   * **Two-sided model (if supported)**   + **i.e., CSI compression (if supported)**   **Note: whether two-sided model is supported or not is separate discussion.**  **Note: whether the model identification is applicable for all inter-vendor training collaboration options is separately discussed.** |
| TCL | Agree |
| ETRI | We support |
| Panasonic | Support  In CSI compression agenda following was agreed. This is not specific to CSI compression but applicable to all AI/ML use case. Regardless of the same environment or condition, when UE physical models are different, it can happen that suddenly expected behaviour is not available. In order to prevent such situation, physical model can be required to be identified by the NW.  Agreement  For option 1 / 3 / 4 / 5 and their sub-options, study mechanisms (e.g., post-deployment performance monitoring) for identifying the cause (e.g., NW side, UE side, data drift) of the performance degradation to guarantee good performance in the field. |
| Spreadtrum | Support |
| CATT, CICTCI | Intuitively, functionality-based LCM is difficult to do model pairing. This conclusion may be correct from high-level.  But one question: does ‘model identification’ here mean a superset for all MI-Options? We do not think MI-Option1 is needed in two-sided model use case. |
| ZTE | OK |
| Samsung | It is better to discuss this after the discussion on the options for the standardization based solution for two-sided model development in 9.1.3.2. As an example, the provision of associated ID for NW-first training along dataset exchange (Option 4) and reference model exchange (Option 3/5) can handle the model pairing problem. Note that exchange of associated ID is not model identification by itself. |
| Qualcomm | support |
| Xiaomi | Support |
| Ericsson | Our understanding is that model identification is not needed in case a reference model is introduced, hence suggesting  **Conclusion:**  **From RAN1 perspective, model identification is needed and beneficial at least for the following case**   * **Two-sided model in case a single reference model is not feasible to specify, (if supported)**   + **i.e., CSI compression (if supported)**   **Note: whether two-sided model is supported or not is separate discussion.** |
| CMCC | Support |
| Fujitsu | Support |
| NEC | Support |
| Nokia | OK with the direction. |
| LG | OK |
| Futurewei | Support |

#### **Proposal 2.1.6**

There is a limited number of company(ies) in favor of IM-Option5. In contrast, based the submitted tdocs and previous discussion, most companies don’t support IM-Option5. Thus, the following proposal is suggested with the aim to focus on study of some other high-priority option(s).

**Proposal 2.1.6**

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | OK |
| Huawei, HiSilicon | Support. |
| vivo | Support |
| TCL | Agree |
| ETRI | We support |
| Panasonic | Support |
| Spreadtrum | Support |
| ZTE | OK |
| Samsung | Ok |
| Qualcomm | OK |
| Xiaomi | Support |
| Ericsson | Support |
| CMCC | Support |
| NEC | Support |
| Nokia | OK |
| LG | Typo: ‘IM’🡪’MI’ |
| Futurewei | Support |

#### **Proposal 2.1.7**

Based on the tdocs, some companies think that even if MI-Option4 is not supported, the UE and NW can achieve the same understanding on the AI model(s). For this case, UE only needs to report its supported AI model(s), which can be done by UE capability signaling. Thus, the following proposal is suggested for further discussion.

**Proposal 2.1.7**

**The model identification procedure dedicated to MI-Option4 is not pursued for Rel-19 normative work.**

**Note: Whether the fully standardized reference model (structure + parameters) can be used for real deployment and the potential spec impact (e.g., UE reports to network which reference AI model(s) is supported via UE capability signaling) are separate discussions (e.g., to be discussed in RAN1 CSI compression session / RAN4)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Huawei, HiSilicon | Support. |
| TCL | Agree |
| Panasonic | Support |
| ZTE | We understand the intention of this proposal, but this proposal itself seems to be confusing.  Maybe we can update it as following.  **Proposal 2.1.7**  **The model identification procedure of MI-Option4 can be realized via model ID report for the reference model(s).** |
| Samsung | Do not agree. This could only be decided after the standardization-based solution for two-sided model training in 9.1.3.2. MI-Option4 can identify both fully standardized model or standardized model structure. |
| Qualcomm | We should wait for the conclusion from 9.1.3.2. |
| Xiaomi | Wait for the progress in section 9.1.3.2 |
| Ericsson | Ok to wait for 9.1.3.2 |
| CMCC | Support |
| Fujitsu | Support |
| NEC | Support |
| Nokia | OK. Note is not needed. |
| LG | OK |
| Futurewei | Support |

#### **Proposal 2.1.8**

Compared to MI-Option1/3, there seems less support of MI-Option 2 for one-sided model. Thus, the following proposal suggested for further discussion.

**Proposal 2.1.8**

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Panasonic | Support |
| Huawei, HiSilicon | If model identification is applied only to two-sided model as proposed in Proposal 2.1.5, both MI-Option 2 and MI-Option 3 are not applicable to one sided model.  **The model identification procedure dedicated to MI-Option2 and MI-Option3 for one-sided model is not pursued for Rel-19 normative work.** |
| vivo | To Huawei, for Option3 we still have model transfer z4 and z2 on table. |
| TCL | Agree |
| Panasonic | OK. |
| Spreadtrum | Support |
| CATT, CICTCI | Fine with FL’s proposal. |
| ZTE | Agree. |
| Qualcomm | OK |
| Xiaomi | OK |
| Ericsson | Support |
| CMCC | Support |
| Fujitsu | Support |
| NEC | Support |
| Nokia | OK |
| Futurewei | Support |

#### **Proposal 2.1.9 (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 2.1.9 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Huawei, HiSilicon | OK |
| vivo | Support |
| TCL | Agree |
|  |  |
|  |  |
|  |  |

# Training data collection for UE-sided model

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Spreadtrum[2] | *Proposal 1: For data collection for UE-side model training, support 1a or we could wait the progress of RAN2.* |
| Google[3] | *Proposal 4: Support the NW and UE to maintain the same understanding on when the UE can perform data collection.* |
| CMCC[5] | *Proposal 15: Regarding the UE side data collection mechanism, RAN2 could take the Reply LS on Data Collection Requirements and Assumptions (R1-2310681) as the baseline.* |
| Intel[6] | *CN/OAM/OTT collection of UE-sided model training data*  *Proposal 14:*   * *On CN/OAM/OTT collection of UE-sided model training data, the benefits from supporting data collection using unspecified format compared to using a standardized data format (that can utilize data collection framework for network-side model training data collection) are unclear.* * *On contents of the collected data, the details listed in R1-2310681 are considered as the baseline inputs from RAN1 for the present study.* |
| ZTE[7] | *Proposal 10: 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.* |
| OPPO[11] | *Proposal 8: On UE data collection, RAN1 waits for RAN2 progress on UE data collection mechanisms based on RAN1’s LS reply in Rel-18 study, and can carry out additional study on if RAN2 needs further assistance.* |
| Xiaomi[12] | *Proposal 9: The data content and related information included in RAN1 LS (R1-2310681) to RAN2 can be set as baseline* |
| Fujitsu[13] | *Proposal-14: From the RAN1 perspective, the data collection for UE-side model training can be further studied in the use case sub agendas.* |
| CATT[14] | *Observation 4: RAN1 already starts the normative work of data collection for UE-side model training within RAN1 scope, including the corresponding contents of UE data collection per WI use case.*  *Proposal 15: For data collection for UE-side model training,*   * *RAN1 focuses on how to collect training data into UE device in air interface, including the corresponding contents of UE data collection per WI use case;* * *RAN2 and higher layers focus on whether and how the training data is transferred/delivered from UE device to UE-side server, e.g. via CN/OAM/OTT.* |
| LG[16] | *Proposal#5. RAN1 to discuss contents for UE-sided model training data collection in each sub-use-case agenda. Thus, no need to discuss in this agenda.* |
| NVIDIA[19] | *Proposal 3: Conclude that there is a need for collection of UE-sided model training data.* |
| Samsung[23] | *Observation#2: 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#3: 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#10: 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.* |
| ETRI[25] | *Proposal 5: Datasets should be categorized based on NW configurations and configured functionalities during the data collection process.*  *Proposal 6: The NW can request UEs to transfer collected data immediately for the purpose of categorizing the dataset.*  *Proposal 7: The UE needs a mechanism to categorize data samples according to changes in its settings.* |
| Apple[26] | *Proposal 6: From RAN1 perspective, option 1-1a is sufficient and no additional requirement is identified to enhance data collection for UE side model training.* |
| DOCOMO[29] | *Proposal 3: When considering data collection toward OTT/UE side server, the ownership of data should be clarified first.* |
| Huawei[31] | *Proposal 12: 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.* |
| Qualcomm[32] | *Proposal 2: The RAN1/RAN2 discussion should be focused on data collection for model training on the UE side, considering the following*   * *Direct transfer of the collected data to the server for data collection for UE side training (in a 3GPP transparent or 3GPP non-transparent method)* * *Transfer of the collected data to the server for data collection for UE side training (via CN or OAM).*   *Observation 1: The actual input/output and side/auxiliary information for a UE-side model are implementations-specific choices and cannot be pre-determined/standardized.*  *Observation 2: The auxiliary/side information collected for the model development can be proprietary. Therefore, the data collected from/by a UE vendor should not be shared with other UE vendors, network vendors, operators (without service level agreement between operators and UE vendors), or third parties.*  *Proposal 3: A data collection method that cannot ensure the protection of the UE proprietary information cannot be used as data collection for UE-sided model training.*  *Observation 3: During the runtime, which model(s) UE can run depends upon several UE conditions, e.g., UE power status, UE memory, the coexistence of different AI/ML features, the coexistence of AI/ML features with non-AI/ML feature, and others.*  *Proposal 4: Considering the implementation-specific nature of the model input/output and auxiliary/side information and considering the runtime constraints (as mentioned in observation 3), the UE-side model can only be trained by the UE vendor, at least in the Rel-19 and foreseeable near future.* |
|  |  |

#### **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.

Based on the tdocs, most companies think that the key information for UE-sided data collection has been contained in Rel-18 LS and RAN1 can do some study per RAN2’s request/LS.

Meanwhile, several companies suggest to further study some issues, e.g., assistance/auxiliary/side information, overhead reduction, time window and so on. However, most of the other issues are only mentioned by 1 or 2 companies, e.g., unspecified format for NW-side model training data collection, quantization, availability/quality of ground-truth label, overhead reduction, down-selection of RAN2 solutions, control of the collected data by NW, prioritization of the entities that collects training data.

For other issues/proposals, moderator feels they are more suitable to be discussed in BM/Positioning session, rather than in this session.

From the proposals of companies, it seems the attitudes toward the UE-side training data collection are still diverging. A few of companies believe current progress is sufficient at least in this agenda item, some think RAN1 can be triggered by RAN2 to continue the study and others believe UE-side training data collection can be studied and specified (if needed) in each sub use cases, since the LS in R1-2310681 is not sufficient (lack of details).

**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] | *Observation 2: Case z1 may be deprioritized for R19 as it does not provide clear benefit over Case y; the major difference is the location of their storage. R19 can start with two options: Case y for 3GPP-transparent scenarios and Case z4 for non-3GPP-transparent scenarios.*  *Proposal 14: Deprioritize Case z1 for R19 as it does not provide clear benefit over Case y.* |
| Spreadtrum[2] | *Proposal 2: From RAN1 perspective, the model transfer/delivery Case z1 is deprioritized in Rel-19.*  *Proposal 3: From RAN1 perspective, the model transfer/delivery Case z2 is deprioritized in Rel-19 for two-sided model.*  *Proposal 4: Suggest to defer the discussion on Case z4 until good progress on AI9.1.3.2 multi-vendor issue achieved.* |
| Tejas[4] | *Proposal 5: Consider Alt1 for model delivery/transfer in Case z4*   * *Alt. A*   + *Step A-1: UE reports the supported known model structure(s) to network*   + *Step A-2: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step A-1*   + *Step A-3: based on received parameters, the UE compiles and tests if needed.*   + *FFS: whether some additional step(s), and/or whether other information is needed*   *Proposal 6: For model delivery/transfer Case z4 Alt. A should be prioritised.*  *Proposal 7: For the open format for model delivery/transfer Case z4, reuse the existing open format(s) that is already existing in the AI community (for e.g., ONNX).* |
| CMCC[5] | *Proposal 12: 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 13: It is suggested to further study model transfer/delivery Case z4, from the following aspects:*   * *How to standardize reference model structure* * *How to exchange model parameters* * *The associated procedure*   *Proposal 14: For Alt A in model delivery/transfer Case z4, there could be Step A-0 about the UE capability report on model transfer/delivery case z4:*   * *Alt. A*   + *Step A-0: UE reports to NW its support of model transfer/delivery case z4*   + *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* |
| Intel[6] | *Observation 6:*   * *Collaboration level y offers a basic method for collaboration between UE and the network with limited specification impact but reduced efficiency due to reliance on offline coordination and model delivery as against over-the-air model transfer in use-cases involving site-/scenario-/configuration-specific models.*   *Observation 7:*   * *Offline model compilation and offline model testing, while desirable in general from perspective of UE implementation and model robustness, may not always be essential or justified considering the adverse impact to incurred latency for model updates and/or switching, e.g., for cases wherein model may be updated with respect to limited number of parameters while maintaining the model structure.*   *Proposal 8:*   * *From RAN1 perspective, the model transfer/delivery Case z2 is deprioritized for UE part of two-sided model in Rel-19 due to the following reasons:*   + *Risk of proprietary design disclosure*   + *Burden of offline cross-vendor collaboration*   *Proposal 9:*   * *From RAN1 perspective, model transfer/delivery Case z1 is deprioritized in Rel-19 due to the following reasons:*   + *Not much benefit compared to Case y.*   + *Large burden of offline cross-vendor collaboration.*   + *Additional burden on model storage within in 3GPP network.*   + *Limited applicability to only scenarios involving two-sided models with model transfer/delivery from UE to NW side.*   *Proposal 10:*   * *In Rel-19, consider support of model transfer/delivery Case y and model transfer/delivery Case z4 for model/parameter transfer/delivery.*   + *For model transfer/delivery Case z4, consider specifying a group/family of model structures/backbones to alleviate the burden of offline inter-vendor collaboration to align on model structure between NW and UE.*   *Observation 8:*   * *For model delivery/transfer Case z4, for Alt. B, it is not necessary to include step B-1 as listed* * *For model delivery/transfer Case z4, step B-0 can be interpreted to be implied by the UE reporting in step A-1.*   *Proposal 11:*   * *To support model delivery/transfer Case z4, it is sufficient to follow the procedure as in Alt. A:*   + *Step A-1: UE reports the supported known model structure(s) to network*     - *UE’s support of model transfer/delivery case z4 can be explicitly or implicitly associated with this reporting.*   + *Step A-2: NW transfers to UE the parameters for one or more of supported known model structure(s) reported in Step A-1.*   *Proposal 12:*   * *To support model delivery/transfer Case z4, options to align “known model structure(s)” between NW/NW-side and UE/UE-side include:*   + *Opt. 1: based on specified candidate model structures,*   + *Opt. 2: based on alignment via inter-vendor collaboration,*  |  |  |  | | --- | --- | --- | | *Category* | *Parameter* | *Description/Examples* | | *Model architecture parameters* | *Model type* | *Transformer, CNN, RNN, MLP* | | *Model depth* | *Number of layers* | | *Layer type* | *Fully connected, convolutional, activation layer, etc.* | | *Layer size* | *Neuron count and configuration* | | *Quantization method for the encoder output* | *Scalar, vector (with codebook)* | | *Encoder-decoder interface* | *Number of bits of latent message* | | *Fixed point representation* | *Int8, int16, floating point, etc.* | | *Format of input to encoder/output of decoder* | *Similar to options for training data representation* | | *Compression type* | *SF (spatial-frequency) vs. TSF (time-spatial-frequency) for CSI compression* |  * + *Opt. 3: based on a combination of candidates specified for certain model structure parameters and further alignment based via inter-vendor collaboration.*   *Proposal 13:*   * *To support model delivery/transfer Case z4 for two-sided models for CSI compression, to align “known model structure(s)” between NW/NW-side and UE/UE-side at least candidates for one or more of parameters defining model structure are specified.* * *RAN1 to further consider feasibility of specifying candidate values for one or more of the following parameters:* |
| ZTE[7] | *Proposal 7: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer study for two-sided model rather than UE-side model.*  *Observation#6: the overall prioritization up to RAN1#116bis is of the following.*   |  |  |  | | --- | --- | --- | |  | *UE-sided model* | *Two-sided model* | | *Case y* |  |  | | *Case z1* |  |  | | *Case z2* | *Deprioritized (RAN1#116bis)* |  | | *Case z3* | *Deprioritized (RAN1#116bis)* | *Deprioritized (RAN1#116bis)* | | *Case z4* |  |  | | *Case z5* | *Deprioritized (RAN1#116)* | *Deprioritized (RAN1#116)* |   *Proposal 9: The details of Alt. A and Alt. B of model transfer can be further clarified in normative work phase, e.g., information exchange about the buffered parameters at the UE side.* |
| Continental Automotive[8] | *Proposal 3: Study additional aspects with details about indication of UE capability and the associated signaling overhead for the unified procedure of model transfer/delivery.*  *Proposal 4: Study UE reporting specifics about model information related to the ongoing discussion such as model identification process and model transfer/delivery.*  *Proposal 5: Study of training data quality is suggested within the scope of general framework for RAN1 aspect.* |
| Ericsson[9] | *Proposal 12 Rel-19 RAN groups prioritize case y for model delivery, if a need arises based on use case progress, and down-prioritize the other cases.*  *Proposal 13 Only if the collaboration burden of case y with NW-sided training is deemed infeasible, prioritize case z4 with specified model structure and coefficient precision.*  *Proposal 14 Conclude, from RAN1 perspective, that the model transfer/delivery Case z2 is deprioritized also for UE-part of two-sided model in Rel-19 due to the following reasons: •Risk of proprietary design disclosure •Burden of offline cross-vendor collaboration* |
| vivo[10] | *Proposal 12: RAN4 has an aligned model structure pair for both encoder and decoder of CSI compression for feasibility study purpose, which could be a starting point for specified model structure(s) in RAN1.*  *Proposal 13: The reference model structure may be aligned through the following procedures*   * *Step 0: Align evaluation assumptions* * *Step 1: Determine the model backbone based on consensus and evaluation results on complexity and performance.* * *Step 2: Determine the model hyperparameters that need to be aligned.* * *Step 3: Align the hyperparameters of the model.*   *Proposal 14: 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 8: The burden of model storage would be relieved if the model structure is specified in 3GPP.*  *Observation 9: Proprietary design disclosure may not be a concern if the model structure is specified in 3GPP.*  *Proposal 15: RAN1 can further conclude on feasibility of model parameter update for Case z4 with model structure specified in 3GPP.*  *Proposal 16: RAN1 can further conclude on necessity of model transfer/delivery based on e.g., inter-vendor collaboration needs for two sided models.*  *Proposal 17: The following additional steps can be added for model delivery/transfer Case z4.*   * *NW could indicate to transmit partially model parameters, in or before Step A-2 or Step B-3* * *UE may report whether AI/ML model is ready to be used, after Step A-2 or B-3.*   *Proposal 18: For defining a new open format within 3GPP for model transfer Case z4 with specified model structure(s), the specification effect may be acceptable.*  *Proposal 19: Study pros and cons of the following options for model delivery/transfer Case z4 assuming 3GPP would specify the model structure:*   * *Option 1: Reuse the existing open format(s) that has existed in the AI community (e.g., ONNX);* * *Option 2: Define a new open format within 3GPP;* * *Option 3: Reuse the mechanism defined in SA2 (interoperability token) for aligning model description format.* |
| OPPO[11] | *Proposal 7: 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.* |
| Xiaomi[12] | *Observation 1: For the model trained by 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 by 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 model structure alignment between NW and UE*   * *Option 1: Via offline coordination* * *Option 2: Via specified reference model*   *Observation 5:*   * *For Case z4 with offline coordinated model structure, offline co-ordination effort is required* * *For Case z4 with specified reference model, 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 model structure*  *Proposal 3: The model parameter delivery alternatives for Case Z4 should consider the following steps:*   * *Identify the potential need for the model parameter delivery* * *Confirm UE’s consent on the model parameter delivery* |
| Fujitsu[13] | *Model transfer*  *Proposal-9: Deprioritize Case z2 for the two-sided model to avoid the disclosure of proprietary information on model format.*  *Proposal-10: Deprioritize Case z1 if its benefit over Case y from the location of model storage cannot be justified.*  *Proposal-11, Model/Model structure identification and open format are suggested to be clarified in the further study of Case z4.*  *Proposal-12 Study post-model-transfer performance monitoring/test to guarantee the performance after model transfer/model update in the field.*  *Proposal-13 Model delivery/transfer Case z4 can be further studied for both one-sided and two-sided models in Rel-19 with focus on the following aspects:*   * *Model/model structure/model parameters identification.* * *Open format details.* * *Post-model-transfer/update performance monitoring/test in field.* |
| CATT[14] | *Proposal 11: 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 is assuming Step A-1 is not part of UE capability report.*   *Proposal 12: For Alt.B of model transfer case z4,*   * *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.* * *Step B-0 may be part of the UE capability report.*   *Proposal 13: 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: Exchange model ID;* * *Direction 2: Exchange model structure ID;* * *Direction 3: Exchange model structure described by a known model description format.*   *Proposal 14: In Rel-19, for model transfer/delivery Case z4, conclude that:*   * *Necessity: Not necessary, since some other alternatives can achieve similar goal;* * *Feasibility: Not completely infeasible, but huge difficult/effort is foreseen;* * *Benefit: Benefits are identified on inter-vendor collaboration reduction, cell-specific model optimization, and avoiding training data leakage, etc. But some other alternatives can also achieve similar benefits.* |
| LG[16] | *Proposal#6. Focus on discussing the key challenges of model transfer such as offline cross-vendor collaboration, model storage requirements, and proprietary design disclosure issues, instead of further comparing pros/cons of different model transfer cases.* |
| IIT Kanpur[18] | *Proposal 1: Deprioritize case Z1 for Rel-19 due to the following reasons:*   * *Risk of proprietary design disclosure* * *Burden of offline cross-vendor collaboration.*   *Observation: Reporting of supported known model structure(s) by UE to network for model delivery/transfer Case z4 can have privacy concern about UE exposing it’s details of actual model to be developed.*  *Conclusion: An exchange of known model structure(s) in Model delivery/transfer Case z4 between UE to network can be about Actual UE model to be developed or Reference UE model* |
| NVIDIA[19] | *Proposal 4: Continue to study Cases y, z1 and z4 for transferring/delivering AI/ML model(s).* |
| InterDigital[20] | *Observation 11: In cases where model generalization, model finetuning or model storage/switching is not feasible, model delivery/transfer can be beneficial.*  *Proposal 17: Model transfer for UE-side models with functionality-based LCM is not supported and 3GPP specification transparent model delivery is only considered.* |
| NEC[21] | Observation 6: Supporting model transfer is essential when considering cell/scenario-specific AI/ML deployment which is expected to happen when AI/ML deployment accelerates.  *Proposal 15: Model transfer should be supported from Rel-19 to ensure future-proofness of AI/ML operation.*  *Proposal 16: Support specification of model transfer methodology z4.*  *Proposal 17: 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* |
| Nokia[22] | *Observation 1: For two-sided CSI compression, particularly training type 1 (joint model training and model transfer/delivery to the UE), model transfer to be realized as user plane data transfer, controlled by the gNB/RAN using the control plane signaling.*  *Observation 2: 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.* |
| Samsung[23] | *Proposal#11: Deprioritize study on Case z1 of 3GPP non-transparent model transfer cases as it requires offline cross-vendor collaboration.*  *Observation#4: 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#12: Study the feasibility and potential benefits of model (parameter) transfer for specified model structure from gNB to UE, i.e., Case z4.*  *Observation#5 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#13 For model delivery/transfer Case z4 with specified model structure, further study the necessity of model identification starting from MI-Option4.* |
| Panasonic[24] | *Proposal 5: RAN1 should focus Alt. A, i.e. the supported model structure is directly informed to NW in model delivery/transfer Case z4.* |
| Apple[26] | *Proposal 5: For model transfer z4, additional steps and information are added for Alt A and Alt B, including:*   * *UE indicating whether model can be used directly for inference, together with the supported known model structure to NW.* * *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.* |
| AT&T[27] | *Proposal 6: 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.*  *Observation 1: There are benefits and challenges to both proprietary and open format model transfer. It is beneficial to have both specified to support different use cases based on requirements.*  *Proposal 7: Study and specify both proprietary and open format model transfer for both UE-sided models and UE-part of two-sided models in Rel-19.*  *Proposal 8: For model delivery/transfer to UE, from the device implementation point of view*   * *Model delivery/transfer to UE in a proprietary format (Case y, z1) is feasible from the device implementation point of view from RAN1 perspective.* * *Parameter update of a known structure on a deployed model via model delivery/transfer in an open format (Case z4) may be beneficial for certain use cases or deployment scenarios, e.g., when it is desired to have shorter model parameter update timescale due to no need for offline compiling with less offline engineering, but it comes with potential requirements/challenges, e.g., advanced device implementation, lack of device-specific optimization/testing compared to model delivery via proprietary format.*   *Proposal 9: Regarding model transfer/delivery Case z4 for one-sided model, Rel-19 study focuses on the option with standardized known model structure(s).* |
| Meta[28] | 1. *Further discuss model transfer/delivery with collaboration level z4 where UE can indicate support of a subset of known candidate model structures based on UE capability* |
| DOCOMO[29] | *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 6: Discuss how many model structures should be standardized for model transfer case z4.*  *Proposal 7: Discuss the approaches to determine which model structures should be standardized. One practical approach is simulation evaluation with the calibration over companies.*  *Proposal 8: Study the pros and cons of using the existing model format (e.g., ONNX) or introducing new 3GPP format* |
| Huawei[31] | *Observation 4: 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 5: 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 11: For model transfer/delivery where the model is trained at UE side or neutral site, assume Case y as the baseline.* |
| Qualcomm[32] | *Proposal 5: Conclude that the progress made so far on deprioritizing some model transfer/delivery cases is sufficient, and there is no need to further deprioritize the remaining cases.* |
|  |  |

#### **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 |  |
| Case z3 | Deprioritized | Deprioritized |
| Case z4 |  |  |
| Case z5 | Deprioritized | Deprioritized |

## 1st round discussion

#### **Proposal 4.1.1**

In the tdocs, some companies discussed what aspects should be defined for the “known model structures”. Thus, the following proposal is suggested for further discussion:

**Proposed 4.1.1**

**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)**
* **Number of layers**
* **Layer types/structure (e.g., full connected, activation layer and so on)**
* **Layer size**
* **Connected between different layers**
* **Interface of the model input**
* **Fixed point representation (e.g., floating point, Int16 and so on)**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Huawei, HiSilicon | A clarification question: whether the intention is to determine the specified structure, or to determine the content/interpretation of the delivered information? It seems the intention is the former, i.e., RAN1 to study the exact structure from the following aspects?  Another comment: the interface should also include output. In addition, we need to consider the scalability aspect (may be reflected to model design, or pre/post processing).  **Interface of the model input/output, considering the scalability** |
| vivo | We are fine with the direction.  We are also fine with the wording update from Huawei.  Additionally, the known model structure also needs to align on “activation function”. |
| Panasonic | Support |
| Spreadtrum | One question for clarification:   * Does the fourth bullet layer size mean the number of parameters per layer?   One comment:  Regarding to model structure specification, it seems that agenda 9.1.3.2 (Proposal 23a) is also intending to discuss it. We understand 9.1.3.2 is mainly for two-sided use case. However, for one-sided use case, the necessity of supporting case z4 has not been confirmed, and it may be not proper to discuss the details now for one-sided use case. Thus, to avoid parallel discussions, we think the issue can be delayed for the discussion in this agenda. |
| CATT, CICTCI | Maybe OK, but just afraid that completing the whole list will be time consuming....  Miner suggestion:   * Connection between different layers * Interface of the model input/output   From vivo’s comment, indeed activation function is important for understand structure. It may or may not be part of ‘Connected between different layers’ – better to clarity. |
| ZTE | We are generally fine to go with direction. Some detailed comments from our side.  Firstly, the model structure can also be applied for reference model discussed in RAN1.  Secondly, it seems that this list isn’t fully aligned with RAN4’s outcome. |
| Qualcomm | The intention of the proposal is not clear. Why do we need to list all the items for known model structure while the purpose is not to specify such model structure. |
| Ericsson | Although there is a large momentum around Neural Networks, it is unclear how RAN1 can make agreements where we limit the “known model structures” to neural networks (e.g. by defining Layer size). Decision trees could be another useful model architecture.  **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**   + **Connected between different layers** * **Interface of the model input** * **Fixed point representation (e.g., floating point, Int16 and so on)** |
| Fujitsu | For model transfer aspect, we don’t say to discuss the details on the known model structure. The known model structure can be indicated via model structure ID or model ID. The specifying model structure could be a separate topic if needed. |
| Nokia | Ok with the direction |
| LG | Similar view with Spreadtrum. This is mainly for two-sided model, thus this detail can be discussed in CSI agenda. |
| Futurewei | OK as a starting point. Ericsson’s revision is also supported. |

#### **Proposal 4.1.2**

In additional to the two-sided model, some companies believe Case z4 is applicable to the scenarios with cell-specific model (e.g., local AI models). For the two-sided model, we can discuss it later so that we can ensure the alignment with CSI compression session. Thus, we focus the discussion on one-sided model here.

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

* 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
* …

In order to reduce the workload and be aligned with the discussion of CSI compression, the following proposal is suggested to restrict the scope of Case z4.

**Proposal 4.1.2**

**Regarding the study of necessity/benefit of model transfer/delivery Case z4 for one-sided model, Rel-19 study focuses on the option with standardized known model structure(s).**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| NTT DOCOMO | Fine with the proposal. |
| Huawei, HiSilicon | This proposal conflicts with Proposal 2.1.5. If model identification is only applicable to two sided model, then model transfer/delivery is not applicable to one-sided model.  Suggest to deferring this proposal until we have conclusion on Proposal 2.1.5. Or, we remove “one-sided model” from this proposal to make it more generic.  **Regarding the study of necessity/benefit of model transfer/delivery Case z4 ~~for one-sided model~~, Rel-19 study focuses on the option with standardized known model structure(s).** |
| vivo | Fine with this proposal or with the general wording from HW. |
| Panasonic | Support |
| Spreadtrum | Fine with update from HW |
| CATT, CICTCI | Support. |
| ZTE | We believe the case z4 discussion should prioritize two-sided model. If it is deemed that case z4 for two-sided model is necessary, we can further clarify whether it can be applied to UE-sided model. |
| Samsung | Support. |
| Qualcomm | This has serious feasibility concerns. Let us consider the beam prediction use case. Across different NW vendors the NW-side additional conditions are quite different (including at least the size of Set A, and the relationships between Set A, Set B beams). Also, the implications for different use cases (and different sub-use cases are quite different). Example: for temporal beam prediction: how far into the future we predict can be part of the model design. How can we pre-define different model structures which satisfy the requirements of different use cases? As another example, let us consider spatial beam prediction. When Set B is not a subset of Set A and when Set B is a subset of Set A. These are quite different use cases with different AI/ML model requirements. Depending on at least the size and relative beam shape relationship of Set A and Set B beams we may need more or less sophisticated AI/ML model structures. So looking from a feasibility perspective, we do not support the proposal. |
| Ericsson | Same view as HW |
| Fujitsu | The benefits and feasibility of standardized known model structure(s) for one-sided model need to be clarified. |
| NEC | Support |
| Nokia | For one sided cases, there is no clear reason to support this. |
| LG | Huawei’s version seems ok |
| Futurewei | Support HW’s revision. |

#### **Proposal 4.1.3**

In order to assess the feasibility/benefit/spec impact of Case z4, the specification efforts on the open format should also be considered. Thus, the following proposal is suggested for discussion:

**Proposal 4.1.3**

**For the open format for model delivery/transfer Case z4, further study the following Options (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**
* **Option 3: Using ASN.1 to represent the AI model**
* **Option 4: Reuse the mechanism defined in SA2 (interoperability token) for aligning model description format.**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Huawei, HiSilicon | If we consider Case z4, as the delivered information are only parameters (no need to describe model structure), why do we still need to specify the model representation format (ONNX or other formats)? ASN.1 is enough to deliver the parameters. |
| vivo | To Huawei, we are fine to go directly to Option 3. But I think FL is trying to put everything on table for now. Even without model structure information, we can also use some open formats. |
| Panasonic | Support |
| CATT, CICTCI | OK do discuss. But in general, once model format is known and aligned between NW and UE, we think the formats used to describe model structure are redundant. |
| Qualcomm | Discussion is dependent on Proposal 4.1.2 |
| Ericsson | What is the difference between option 2 and 3 ? Option 2 defines a format that is not based on ASN1? |
| Nokia | Needs some more discussions to understand the options. Some options can be combined (option 2 and 3) as we understand. |
| LG | This topic may be better fit to RAN2 or RAN3 |
| Futurewei | OK to study. |

#### **Proposal 4.1.4**

In the tdocs, most companies tend to deprioritize Case z1 and z2. In RAN1#116bis meeting, during the online session, QC argued that the spec impact of z1 is not within RAN1 scope and then Vice Chair concluded that z1 will not be discussed further in RAN1. Therefore, in this summary, no proposal is suggested for Case z1.

Regarding Case z2,

* In previous meetings, most companies expect one agreement to deprioritize Case z2 for both two-sided and one-side model
* In the submitted tdocs, most companies continue proposing to deprioritize Case z2 for two-sided model

**Proposal 4.1.4**

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

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

Companies can provide comments/inputs in the following table:

|  |  |  |
| --- | --- | --- |
| Company | Comment | |
| NTT DOCOMO | Support. | |
| Huawei, HiSilicon | Support | |
| Panasonic | Support | |
| Spreadtrum | Support | |
| CATT, CICTCI | Support | |
| ZTE | Agree | |
| Samsung | Support | |
| Qualcomm | As mentioned before, Casez z2 and z4 are related to the ongoing discussions with regards to Option 3 and Option 5 in CSI compression agenda item. Unless we have a conclusion from 9.1.3.2, we should not deprioritize these options. |
| Ericsson | Support | |
| CMCC | Support | |
| Fujitsu | Support | |
| Nokia | OK | |
| LG | OK | |
| Futurewei | Support | |

#### **Proposal 4.1.5 (Placeholder)**

**Proposal 4.1.5 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

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

# Others

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Continental Automotive[8] | *Proposal 6: Support assessment/monitoring of the performance of inactive models at the UE side.*  *Proposal 7: Study of online training is suggested for inclusion in the upcoming additive discussion.* |
| Ericsson[9] | *Observation 11 The study phase of the Rel-19 WI did not bring any new findings or agreements that justify the need to support model-based LCM for one-sided models, beyond the TR conclusions documented during the Rel-18 SI.*  *Proposal 15 Conclude that Functionality based LCM is sufficient for one-sided use cases within Rel-19 scope.*  *Proposal 16 Unless two-sided use case is agreed to be part of Rel-19 normative work, there is no need to study further model identification or model-based LCM in the scope of Rel-19.*  *Proposal 17 Unless two-sided use case based on option 3 (Standardized reference model structure + Parameter exchange between NW-side and UE-side) is agreed to be part of Rel-19 normative work, there is no need to consider model transfer (case z4) in the scope of Rel-19.*  *Proposal 18 Conclude that the Rel-18 LS response to RAN2 is sufficient from a RAN1 perspective for addressing the study objective on data content, there is no further need to consider general UE-side data collection in the scope of Rel-19.* |
| TCL[15] | *Observation 3: The associated ID can be designed according to the additional conditions group to reduce the overhead and provide clear mapping to physical parameters.* |
| NVIDIA[19] | *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.* |
| InterDigital[20] | *Observation 6: A ground truth label quality indicator generated by a UE or PRU may be unreliable as the estimate UE location may be inaccurate*  *Observation 7: For UE side model, additional specification impact for UE reporting is not needed, but a procedure to measure whole Set A over multiple time instances is needed.*  *Observation 8: For gNB side model, enhancement of UE reporting is needed as gNB needs to acquire UE side measurements.*  *Observation 9: Compared to data collection for inference, data collection for training requires huge overhead for both BM-Case 1 and BM-Case 2.*  *Observation 10: According to the evaluation results, measured RSRPs within one UE do not significantly change over different beams in spatial domain and different time instances within one beam.*  *Proposal 7: For model input and ground truth for CSI prediction model training dataset, the collected data could include the measured CSI during the observation and the prediction window.*  *Proposal 8: Other information for the CSI prediction model training dataset could include the sizes of the observation and prediction windows, CSI format (raw or eigenvector), pre-processing (if any), CSI-RS configuration, the number of Tx antenna ports and BWP and sub-size.*  *Proposal 9: Quality indicators for the CSI prediction model training dataset could include at least the RSRP and TDCP.*  *Proposal 10: A ground truth label quality indicator is associated with a UE or PRU location*  *Proposal 11: For case 1 for positioning, support LMF to forward location information of PRUs, measurements made by PRUs and ground truth label quality indicator with the PRU location to a target UE*  *Proposal 12: For case 1 for positioning, support LMF to forward location information of a UE, which is not a PRU, measurements made by the UE and ground truth label quality indicator associated with the UE location to a target UE*  *Proposal 13: The LMF is the only entity that can generate a ground truth label quality indicator associated with location information of UE or PRU*  *Proposal 14: For UE side model, support a common procedure to measure whole Set A over multiple time instances for both BM-Case 1 and BM-Case 2.*  *Proposal 15: For gNB side model, support enhanced UE reporting to report up to 64 RSRP values for whole Set A over multiple time instances.*   * *No CRIs/SSBRIs are reported and implicit beam indexes (e.g., by association with RSs and reported RSRPs) are used.* * *Information on measured past instances (e.g., time stamp) is supported.*   *Proposal 16: Support beam reporting compression mechanism for training to reduce overhead by using RSRPs in neighboring beams in spatial domain and RSRPs within a same beam in temporal domain.* |
| NEC[21] | *Observation 7: 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 8: 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 18: 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.* |
| AT&T[27] | *Proposal 10: The following table captures the different approaches through which the additional conditions can be indicated and how they can provide the consistency between the training and inference.*   |  |  |  |  | | --- | --- | --- | --- | | *Approach* | *How NW-side additional conditions are indicated* | *How to ensure consistency between training and inference regarding NW-side additional conditions* | *Analysis* | | *Model identification Type A* | *Aligned offline* | *Indicated via an ID (model ID or ID for additional condition) for model selection* | *There is an offline alignment between the NW and UE regarding additional conditions and the associated model ID. The NW provides the model ID for the correct model to select for the UE based on its additional conditions.* | | *Model Identification Type B2/ Model training at NW and transfer to UE* | *NW provides an ID in form of dataset ID or part of model ID to the UE. The UE reports the model ID for the model trained using these additional conditions.* | *The NW provides the UE with the ID for model selection* | *The NW provides an ID such as dataset ID or model ID (or part of model ID). The UE provides/confirms the model ID that was trained for the additional conditions. The NW can provide the model ID to select the appropriate model at the UE.* | | *Assistance information* | *Provided to UE for dataset categorization in the form of an ID (determined by the NW)* | *Provided to UE for (transparent) model selection in the form of ID* | *The NW generates an ID for its additional conditions for data collection and provides it to UE to train appropriate models. The NW can later provide the additional condition during inference to assist the UE to transparently select the appropriate model.* | | *Assisted Monitoring* | *NW provides an ID for additional condition to the UE* | *..* | *For the models at the UE the NW provides an ID for the additional conditions. It can be provide assistance to the UE to determine if switch or turn off its model for certain additional condition (as performance requirements would not be met).* |   *Proposal 11: For inference for UE-side models, to ensure consistency between training and inference regarding UE-side additional conditions (if identified), the following options can be taken as potential approaches (when feasible and necessary):*   * *UE handles UE-side additional conditions transparently to NW.* * *Model identification to achieve alignment on the UE-side additional condition between NW-side and UE-side* * *Information and/or indication on UE-side additional conditions is provided to NW.* * *Consistency assisted by monitoring (by UE and/or NW, the performance of UE-side candidate models/functionalities to select a model/functionality)* * *UE report/update of applicable model/functionality based on UE-side additional condition.* * *Other approaches are not precluded.* * *Note: it does not deny the possibility that different approaches can achieve the same function.*   *Proposal 12: For inference for NW-side models, to ensure consistency between training and inference regarding UE-side additional conditions (if identified), the following options can be taken as potential approaches (when feasible and necessary):*   * *Alignment on the UE-side additional condition between NW-side and UE-side* * *Information and/or indication on UE-side additional conditions is provided to NW* * *Consistency assisted by monitoring (by UE and/or NW)* * *Other approaches are not precluded,*   *Note: it does not deny the possibility that different approaches can achieve the same function.*  *Proposal 13: For inference for two-sided models, to ensure consistency between training and inference regarding NW-side and UE-side additional conditions (if identified), the following options can be taken as potential approaches (when feasible and necessary):*   * *Pairing establishment (i.e., model identification) to achieve alignment on the additional conditions between NW-side and UE-side* * *Model training at NW and transfer to UE, where the model has been trained under the NW-side additional condition.*   + *FFS: How to address UE-side additional conditions (if necessary)* * *Other approaches are not precluded.*   *Note: it does not deny the possibility that different approaches can achieve the same function.* |

**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 online session

# 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.**

# 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](mailto:guan_peng@nec.cn)  [pravjyot.deogun@EMEA.NEC.COM](mailto: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](mailto: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. R1-2405810 Discussion on other aspects of AI/ML model and data on AI/ML for NR air-interface FUTUREWEI
2. R1-2405902 Discussion on other aspects of AI/ML model and data Spreadtrum Communications
3. R1-2405954 AI/ML Model and Data Google
4. R1-2405962 Other aspects of AI/ML Model and Data Tejas Networks Limited
5. R1-2405979 Discussion on other aspects of AI/ML model and data CMCC
6. R1-2406018 Other study aspects of AI/ML for air interface Intel Corporation
7. R1-2406058 Discussion on other aspects of AI/ML model and data ZTE Corporation, Sanechips
8. R1-2406064 Discussion on other aspects of AI/ML model and data Continental Automotive
9. R1-2406142 Discussion on other aspects of AI/ML Ericsson
10. R1-2406176 Other aspects of AI/ML model and data vivo
11. R1-2406258 Additional study on other aspects of AI/ML model and data OPPO
12. R1-2406273 Further study on AI/ML model and data Xiaomi
13. R1-2406309 Discussion on other aspects of AI/ML model and data Fujitsu
14. R1-2406357 Further study on AI/ML for other aspects CATT, CICTCI
15. R1-2406397 Discussions on Other Aspects of AIML In NR Airinterface TCL
16. R1-2406419 Discussion on other aspects of AI/ML model and data LG Electronics
17. R1-2406444 Discussion on other aspects of AI/ML model and data Lenovo
18. R1-2406459 Discussion on other aspects of AI/ML model and data IIT Kanpur
19. R1-2406496 Additional study on other aspects of AI model and data NVIDIA
20. R1-2406503 Discussion on other aspects of AI/ML model and data InterDigital, Inc.
21. R1-2406542 Discussion on other aspects of AI/ML model and data NEC
22. R1-2406590 Other Aspects of AI/ML Model and Data Nokia
23. R1-2406641 Discussion for further study on other aspects of AI/ML model and data Samsung
24. R1-2406674 Discussion on other aspects for AI/ML for air interface Panasonic
25. R1-2406721 Discussion on other aspects of AI/ML model and data ETRI
26. R1-2406830 Discussion on other aspects of AI/ML model and data Apple
27. R1-2406872 Other Aspects of AI/ML framework AT&T
28. R1-2406889 Other Aspects of AI/ML Model and Data Meta Ireland
29. R1-2406924 Discussion on other aspects of AI/ML model and data NTT DOCOMO, INC.
30. R1-2406964 Discussion on other aspects of AI/ML model and data Sharp
31. R1-2406976 Discussion on other aspects of the additional study for AI/ML Huawei, HiSilicon
32. R1-2407023 Other aspects of AI/ML model and data Qualcomm Incorporated