**3GPP TSG RAN WG1 #116-bis R1-240XXXX**

**Changsha, Hunan Province, China, April 15th – 19th, 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:

|  |  |
| --- | --- |
| Huawei[1] | *Observation 1: The boundary between model identification and functionality identification for the Functionality with model ID is not clear.*  *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.*  *Observation 2: Model-ID-based identification/LCM is applied with globally unique model ID. As a difference, functionality-based identification/LCM is not applied with globally unique ID.*  *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: For two-sided model, the ID-related information has more specific meaning for model identification in model training and inference phase than data collection related configurations, and whether the model ID is needed during the data collection phase can be further studied.*  *Observation 3: For MI-Option 1, 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.*  *Proposal 4: MI-Option 1 is not applicable to one-sided model case.*  *Proposal 5: Data categorization indication for UE side data collection of one-sided model, if needed, could be studied with local ID rather than globally unique ID.*  *Proposal 6: MI-Option 2 is applicable to two-sided model case.*  *Proposal 7: 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 8: 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 9: MI-Option 3 is applicable to two-sided model case.*  *Proposal 10: 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 11: For the procedure of MI-Option 3, the model identification is achieved when the model ID is delivered in together with the delivered model.*  *Observation 4: 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.* |
| FUTUREWEI[2] | *Proposal 1: Do not support MI-Option 2 unless the following are clarified.*   * *The boundary between MI-Option 1 (MI with data collection related configuration(s) and/or indication(s)) and MI-Option 2 (MI with dataset transfer), as both options are related to data collection/the 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 2: Study the following, if the issues in Proposal 1 have been clarified and MI-Option 2 is supported.*   * *Method of referring to a dataset.* * *Necessity of dataset transfer and the mechanism of doing it.*   *Proposal 3: Support MI-Option 3 with further study of its procedures and other details, once the mechanism of model transfer is determined (standardized or non-standardized).*  *Proposal 4: Proponents to justify MI-Option 4 as one of the valid options for model identification.*  *Proposal 5: For MI-Option 4, if justified, clarify how to identify a model through standardized reference model, from the following aspects.*   * *The relationship between the reference model and multiple derived models, in the case only the structure of the reference model is standardized.* * *The level the reference model needs to be standardized, both structure and parameter, or structure-only.*   *Proposal 6: Proponents to justify MI-Option 5 as one of the valid options for model identification.*  *Proposal 7: For MI-Option 5, if justified, clarify the procedure of identifying a model via model monitoring.*  *Proposal 8: For MI-Option 5, if justified, clarify whether MI-Option 5 (Model identification via model monitoring) requires performance monitoring of unidentified and inactive models.*  *Proposal 9: For MI-Option 1, clarify the relationship between model ID and the corresponding configuration for data collection, in the case multiple models can be developed/trained using the same dataset.*  *Proposal 10: For MI-Option 1, conclude that model IDs are assigned only by the NW.* |
| Ericsson[3] | *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.*  *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 Conclude that how information and/or indication on NW-side additional conditions is used by the UE to ensure consistency between training and inference is transparent to the NW.*  *Proposal 3 Conclude that it is not necessary to assign model IDs to UE-sided models.*  *Observation 2 The applicability signalling of a functionality in a reactive approach mitigates the need for model-ID based LCM.*  *Observation 3 During training, Model identification can be implicitly initiated by the NW by indicating the NW additional conditions that is used to ensure consistency between training and inference. Afterwards, UE identifies its models by reporting applicability of the supported AI-based functionality. No explicit Model ID is needed.*  *Proposal 4 For MI-Option 1, for one-sided models, consider the following new information (needs standardization) for “data collection related configuration(s) and/or indication(s))”, which is transmitted from NW to UE,*  *o Beam management: consistency identifiers corresponding to whether the UE can assume during inference that the CSI resources are using the same NW spatial TX-filter (beam/precoders) as during training, and configuration of set A/B,*  *o Positioning: area identifiers corresponding to TRPs transmitting the DL PRS, and selected PRS configuration parameters*  *o CSI prediction: No new information needing standardization identified yet,*  *Proposal 5 For MI-Option 1, for one-sided models, consider the following existing information for “data collection related configuration(s) and/or indication(s))”, which is transmitted from NW to UE,*  *o Beam management: selected SSB/CSI-RS configuration parameters,*  *o Positioning: selected PRS configuration parameters,*  *o CSI prediction: selected CSI-RS Resource/reporting configuration parameters*  *Proposal 6 For MI-Option 1, for the one-sided use cases, regarding “Relationship between model ID and data collection related configuration(s) and/or indication(s)”, conclude that this relationship is a UE internal process and how the UE maps the configuration to a possible model ID is transparent to the NW. There is no need to define a model ID visible to 3GPP specification.*  *Observation 4 For MI-Option 1, for the one-sided use cases, UE can report the applicability of a certain NW configuration (i.e., reactive approach), or a preferred NW configuration (i.e., proactive approach). However, there is no need to introduce a “model ID” to support such applicability process.*  *Proposal 7 For MI-Option 1, conclude that MI-Option 1 is applicable to use cases with UE-sided model. On the other hand, model ID is not necessary for the one-sided use cases.*  *Proposal 8 For MI-Option 1, further study its applicability to the two-sided use case.*  *Observation 5 RAN1 has yet not concluded on the support of training collaboration type 3 and therefore, there is still uncertainty in the need to support Model identification with dataset transfer.*  *Observation 6 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 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.*  *Observation 7 Using ML-option5, it is not clear which possible NW changes that might impact the performance of the UE model. Additionally, significant burden is added to the NW to store history of network configurations including timestamps.*  *Proposal 11 MI-Option 5 should not be considered further.* |
| Spreadtrum[4] | *Proposal 3: For MI-Option 1 and MI-Option 2, one model ID can be mapped into one or multiple dataset ID or dataset collection related configurations.*  *Proposal 4: At least for two-sided model, model identification can be considered, for the sake of providing pairing of two-sided models*  *・ Model-ID-based LCM can be considered and provides more granular, model-level management by NW* |
| Intel[5] | *Observation 1:*  *• Model-ID-based identification is a necessary component to support:*  *o Model transfer from network to UE.*  *o 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.*  *o Model-ID-based identification can apply for all three model identification options (MI-Options 1, 2, 3) subject to support dataset transfer and model transfer for MI-Options 2 and 3 respectively.*  *Proposal 2:*  *• For MI-Option 1, NW may provide a UE with configuration(s) and/or indication(s) for data collection that can be associated with one or more AI/ML model(s) that may be already identified via prior assignment of model ID(s) or be assigned with ID(s) at the time of association to the configuration(s) and/or indication(s).*  *o Alternatively, the model(s) may be identified by associating to provided configuration(s) and/or indication(s) for data collection that, in turn, may be referred to via one or more identifiers provided by the NW.*  *• For MI-Option 1, model IDs can be logical, i.e., multiple physical models may be associated with a set of configuration(s) and/or indication(s) for data collection and share a common model ID.*  *• 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 3:*  *• For MI-Option 2, model identification can be realized via indication of dataset(s) associated with an identified functionality as part of functionality-based LCM or an identified model as part of model-based LCM. The indication of dataset(s) could involve dataset transfer or indication of a previously identified dataset. For both cases, the dataset(s) could be provided with identifiable dataset ID(s).*  *• Model identification may be realized if model-to-dataset mapping is aligned between the UE and the NW, either explicitly or implicitly. Details FFS.*  *• MI-Option 2 can be applicable and beneficial for:*  *o two-sided models for CSI compression use-case,*  *o 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),*  *o localized (site-/cell-specific) models trained at UE-side (or UE-side OTT server).*  *Proposal 4:*  *• 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.*  *o 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 2 can be applicable and beneficial for:*  *o two-sided models for CSI compression use-case,*  *o UE-sided model for which the model is trained at the network side,*  *o localized (site-/cell-specific) models trained at network side.* |
| vivo[6] | *Observation 1: Dataset categorization (or data categorization ID) and model ID have different underlying logic.*  * Dataset categorization 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 dataset categorization.*  *Observation 2: Directly using data categorization information as model ID is not future-proof for cases where real model-level awareness is needed.*  *Proposal 1: To address the issue of maintaining consistency between training and inference, dataset indication/categorization information can be used, rather than model ID.*  *Proposal 2: The main procedure of dataset categorization, is listed as:*  * Step 1: Dataset categorization is assigned by NW during data collection.*  * Step 2: The UE-side model is trained or fine-tuned using the collected dataset and associated dataset categorization information.*  * Step 3: During the inference stage, the current dataset categorization is provided by NW. UE can choose the applicable model based on the information.*  *Observation 3: ID of transferred dataset (if feasible) is not the same as the ID for model identification based on similar reasons as above for data categorization.*  *Observation 4: Feasibility of model identification with dataset transfer is dependent on the feasibility of dataset transfer itself.*  *Proposal 3: Model identification is needed for cases where multiple models are transferred from NW to UE.*  *Proposal 4: 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 5: How model identification via model monitoring works is not clear.* |
| ZTE[7] | *Proposal 2: In Rel-19 AI/ML framework study, type B model identification is prioritized compared with type A model identification.*  *Proposal 3: 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 4: 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.*  *Observation 1: Regarding MI-Option 1,*  * It can only be applied to UE-sided model, but not for two-sided model.*  * The model ID (or additional condition ID) is for each set of data collection related configuration(s) and/or indication(s).*  *Observation 2: Regarding MI-Option 2,*  * It can be applied to UE-side model and two-sided model.*  * The model ID (or dataset ID) is associated with the dataset transferred from base station to UE*  *Observation 3: Regarding MI-Option 3,*  * It can be applied to UE-side model and two-sided model.*  * The model ID is associated with the model transferred from base station to UE*  *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.* |
| Google[8] | *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.* |
| OPPO[9] | *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.*  *• An ID can be used on top of functionality for indication of different additional conditions, to support multiple scenarios, configurations, sites, etc. The ID can be named Model ID or some other name.*  *Proposal 2:*   * *For model identification type A,* * *An ID is allocated to the model as well as the additional conditions used to train the model via OTT inter-vendor engineering.* * *FFS the name of the ID (e.g. model ID, dataset ID, additional condition ID).* * *For model identification type B,* * *MI-Option 1: The gNB can allocate and send an ID corresponding to the model as well as the additional conditions in the training procedure or before the inference starts.*    + *For a UE involved in the model training procedure, gNB can send the ID to the UE in the training procedure.*   + *For a UE not involved in the model training procedure, gNB can send the ID together with the information of the corresponding additional conditions to the UE before the inference starts.* * *MI-Option 2: The gNB can allocate and send an ID corresponding to the dataset as well as the additional conditions together with the dataset transfer in the training procedure.*    + *This option assumes the UE was involved in the model training procedure.* * *MI-Option 3: The gNB can allocate and send an ID corresponding to the model as well as the additional conditions together with the model transfer in the training procedure.*    + *This option assumes the UE was not involved in the model training procedure.* * *Study on MI-Option 4 and MI-Option 5 are deprioritized.* * *FFS the name of the ID (e.g. model ID, dataset ID, additional condition ID).*   *Proposal 3: Functionality ID can be used for indication functionality between NW and UE.*  *Proposal 4: At least after Model identification, 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.*  *Proposal 5: The AI/ML functionality identification, configuration and activation procedure can be as below:*   * *(1) Potential AI/ML functionalities supported by NW and UE are identified based on UE’s and NW’s static capabilities;* * *(2) UE updates the UE capability, and forms the applicable functionality list (which is the sub-set of identified functionality list);* * *(3) NW configures a functionality list, which is a sub-set of applicable functionalities, according to the NW’s instantaneous interest or capability;* * *(4) NW activates a functionality from the configured functionality list.*   *Proposal 6: The AI/ML model identification, configuration and activation procedure can be as below:*   * *(1) Potential AI/ML models supported by NW and UE are identified based on UE’s and NW’s static capabilities;* * *(2) UE updates the UE capability, and forms the applicable model list (which is the sub-set of identified model list);* * *(3) NW configures a model list, which is a sub-set of applicable models, according to the NW’s instantaneous interest or capability;* * *(4) NW activates a model from the configured model list.* |
| CATT[10] | *Observation 1: An AI/ML model can achieve optimal performance when both following conditions are met:*   * *NW-side additional conditions are consistent between the training phase and inference phase;* * *UE-side additional conditions are consistent between the training phase and inference phase.*   *Observation 2: Model identification can provide additional condition information of an AI/ML model, but only in training phase.*  *Observation 3: Model identification itself is not sufficient to support ‘ideal’ model management. In inference phase, additional effort is still needed for the consistency of both NW-side and UE-side additional condition.*   * *If UE takes the control of UE-side model, it needs to know NW-side additional condition of inference phase. The UE does not need standardized model ID to manage UE-sided model.* * *If NW takes the control of UE-side model, it needs to know UE-side additional condition of inference phase. The NW needs standardized model ID to manage UE-sided model.*   *Observation 4: As long as UE-side additional condition is unknown to NW during inference phase, NW cannot make optimum control on UE-side models, unless strong restrictions are set, e.g. the UE-side models always generalize well among all UE-side additional conditions, or the identified UE-side models will always guarantee the consistency of UE-side additional conditions (e.g. UE speed, Rx antenna assumption, sampling frequency error,…).*  *Observation 5: As long as UE-side additional condition is unknown to NW during inference phase, it is better for UE to make decision on the actual model management of UE-side models.*  *Observation 6: Infinite aspects can be categorized as additional condition for signaling. It is unrealistic to sort, list and document all additional conditions for signaling for perfect consistency/alignment.*  *Observation 7: Generalization capability can be one solution to address/alleviate additional condition consistency issue and provide minimum guaranteed performance.*  *Observation 8: Performance monitoring can be another solution to address/alleviate additional condition consistency issue and provide minimum guaranteed performance.*  *Proposal 1: Offline model identification, i.e. type A, is out of 3GPP and cannot be justified by RAN1.*  *Proposal 2: For MI-Option 1, depending on who controls UE-sided model during inference phase, there are two alternatives:*   * *Alt.1, NW controls UE-sided model*   + *Step 1, training data collection phase, NW indicates an NW ID#1 to UE, representing NW-side additional condition;*   + *Step 2, training phase, no 3GPP signaling impact;*   + *Step 3, inference phase, UE reports the NW ID#1 in Step 1 to current NW, in a way of model ID, or carried in meta info, or other ways. NW controls LCM of UE-sided model via model ID, based on NW ID#1 and current NW additional condition.* * *Alt.2, UE controls UE-sided model*   + *Step 1, training data collection phase, NW indicates an NW ID#1 to UE, representing NW-side additional condition;*   + *Step 2, training phase, no 3GPP signaling impact;*   + *Step 3, inference phase, current NW indicates NW ID#2 to UE, representing NW-side additional condition of current NW. UE controls LCM of UE-sided model without model ID, based on the comparison between NW ID#1 and NW ID#2.*   *Proposal 3: For MI-Option 2, depending on who controls UE-sided model during inference phase, there are two alternatives:*   * *Alt.1, NW controls UE-sided model*   + *Step 1, training data collection phase, NW indicates an NW ID#1 to UE, associated with the transferred dataset, representing NW-side additional condition;*   + *Step 2, training phase, no 3GPP signaling impact;*   + *Step 3, inference phase, UE reports the NW ID#1 in Step 1 to current NW, in a way of model ID, or carried in meta info, or other ways. NW controls LCM of UE-sided model via model ID, based on NW ID#1 and current NW additional condition.* * *Alt.2, UE controls UE-sided model*   + *Step 1, training data collection phase, NW indicates an NW ID#1 to UE, associated with the transferred dataset, representing NW-side additional condition;*   + *Step 2, training phase, no 3GPP signaling impact;*   + *Step 3, inference phase, current NW indicates NW ID#2 to UE, representing NW-side additional condition of current NW (but no need to transfer the dataset). UE controls LCM of UE-sided model without model ID, based on the comparison between NW ID#1 and NW ID#2.*   *Proposal 4: If MI-Option 1 or MI-Option 2 is supported, prefer Alt.2, i.e. UE controls UE-sided model.*   * *Only NW indicated ID is supported. No need to support registration/identification of a trained UE-sided model from UE to NW.*   *Proposal 5: For MI-Option 1 and MI-Option 2, if supported, FFS the effective range of NW indicated ID of data collection configuration/procedure or dataset, i.e.*   * *Per cell;* * *Per cell group;* * *Per NW vendor;* * *Per PLMN;* * *Global.*   *Proposal 6: For MI-Option 3, depending on who indicates model structure, there are two alternatives:*   * *Alt.1, UE indicates the supported model structure(s)*   + *Step 1, UE indicates the supported model structure(s) to NW;*   + *Step 2, NW transfers the model to UE, whose structure is supported in UE’s indication in Step 1.* * *Alt.2, NW indicates the candidate model structure(s)*    + *Step 1, NW indicates the candidate model structure(s) to UE;*   + *Step 2, UE reports to NW which structure(s) is supported, among NW’s candidates in Step 1;*   + *Step 3, NW transfers the model to UE, whose structure is supported in UE’s indication in Step 2.* |
| Samsung[11] | *Proposal#1: For MI-Option 1: model identification with data collection related configuration(s) and/or indication(s), consider the following procedure as a starting point*  *• For data collection: Network provides measurement configurations for data collection with indication(s), in the form of an ID, for NW-side additional condition.*  *• For model training: UE-side uses the ID for dataset categorization to train a model compatible with the indicated NW-side additional condition.*  *• For model inference: For UE’s model selection, network provides configuration for inference with indication, in the form of an ID, for NW-side additional condition.*  *Note: The UE-side vendor may develop a single model compatible to multiple NW-side indications (NSIs).*  *Proposal#2: For MI-Option 2: model identification with dataset transfer, consider the following procedure as a starting point*  *• For dataset transfer: Network provides configurations for dataset transfer with indication(s), in the form of an ID, for NW-side additional condition.*  *• For model training: UE-side uses the ID for dataset categorization to train a model compatible with the indicated NW-side additional condition.*  *• For model inference: For UE’s model selection, network provides configuration for inference with indication, in the form of an ID, for NW-side additional condition.*  *Note: The UE-side vendor may develop a single model compatible to multiple NW-side indications (NSIs)*  *Proposal#3: 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#4: For MI-Option 4: model identification via standardization of reference models consider the following options:*  *• MI-Option 4 Type A: Model-ID identifies a standardized reference model*  *• MI Option 4 Type B1: Model-ID indicates UE’s identified model compatible with one or more standardized reference model*  *Proposal#5: For MI-Option 4: model identification via standardization of reference models, UE may indicate supported AI/ML model IDs for a given AI/ML-enabled Feature/FG in a UE capability report.* |
| CMCC[12] | *Proposal 1: The following aspects could be the starting point when discussing the meta info/* *data collection related configuration(s) and/or indication(s) 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: For MI-Option 1, it may include the following types and corresponding procedure:*  *Type 1:*   * *Step1: NW transmit the data collection related configuration(s) and/or indication(s) to UE.* * *Step2: Then UE use this configuration(s) and/or indication(s) to collect training data to train a model. In this way, the ID associated with data collection related configuration(s) and/or indication(s), e.g., resource ID, report ID, ID carried in DCI, can be seen as a type of model ID. But also, NW can assign a model ID explicitly for the model or data collection related configuration(s) and/or indication(s).* * *Note: Before Step1, UE could also report the required data collection related configuration(s) and/or indication(s).*   *Type 2:*   * *Step1: NW transmit the data collection related configuration(s) and/or indication(s) to UE.* * *Step2: Then UE use this configuration(s) and/or indication(s) to collect inference data as the input of model. In this way, the ID associated with data collection related configuration(s) and/or indication(s), e.g., resource ID, report ID, ID carried in DCI, can be seen as a type of model ID. But also, NW can assign a model ID explicitly for the model or data collection related configuration(s) and/or indication(s).* * *Note: Before Step1, UE could also report the required data collection related configuration(s) and/or indication(s).*   *Type 3:*   * *Step1: UE reports the supported/applicable data collection related configuration(s) and/or indication(s) to NW.* * *Step2: Then NW are aware of the model existence/application based on the configuration(s) and/or indication(s). NW can assign a model ID explicitly for the model or data collection related configuration(s) and/or indication(s).*   *Proposal 3: For MI-Option 2, it may include the following procedure:*   * *Step1: Model information exchange between NW and UE.* * *Step2: NW may transfer dataset and assign the model ID to UE side for the following model deployment, model inference and corresponding LCM operation. Also, the model ID can be dataset ID, or the IDs related with dataset transfer triggering/activation/configuration/indication.*   *Proposal 4: 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.* |
| LGE[13] | *Proposal#3. Clarify that any LCM that does not require assigning model ID belongs to functionality-based LCM.*  *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.*  *o for one-sided model cases, other means to provide information/indication for scenario/site-specific models can be considered under functionality-based LCM framework.* |
| Xiaomi[14] | *Observation 4: Model identification is necessary/beneficial in the following aspects*  *- Network additional condition alignment*  *- Two-sided model pairing*  *- Model transfer*  *- Potential processing interruption management*  *- Reducing network burden in handling the additional condition*  *Proposal 3: Consider the following procedure for Type A model identification*  * Step 1:*  *- Train/Update the AI model offline*  * Step 2:*  *- 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 sides assign the model ID for this model to UE side offline*  * Step 3:*  *- UE reports the model ID to network to indicate the availability of the model*  *Proposal 4: The identifier of the data set or data configuration is not equivalents to the model ID*  *Proposal 5: For MI-Option 1*  *- It can be applied to one-sided model if it is supported*  *- The following procedure can be considered*  *o Step 1: NW configures the data collection at the same time, a dataset ID rather than model ID is assigned for the dataset or UE request the data collection and network assign the dataset ID.*  *o Step 2: UE develops the model based on the collected data. Possibly, UE could develop one AI model based on one dataset or multiple datasets.*  *o Step 3: UE side reports the existence of the model together with the associated with dataset ID and other meta information*  *o Step 4: NW assigns the model ID to the model*  *o Step 5: UE reports the model ID to indicate the availability of the model*  *Proposal 6: For MI-Option 2*  *- It can be applied to both one-sided model and two-sided model, if it is supported*  *- The following procedure can be considered*  *o Step 1: NW transfers the dataset to UE and dataset ID is assigned.*  *o Step 2: UE develops the model based on the collected data. Possibly, UE could develop one AI model based on one dataset or multiple datasets.*  *o Step 3: UE side reports the existence of the model together with the associated with dataset ID and other meta information*  *o Step 4: NW assigns the model ID to the model*  *o Step 5: UE reports the model ID to indicate the availability of the model*  *Proposal 7: 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* |
| Panasonic[15] | *Observation 1: MI-Option 1 is the model is identified by the environment of data set for the training. MI-Option 2 is the model is identified by the data set for the training. MI-Option 3 is the model is identified by actual the model parameters and structure.*  *Observation 2: MI-Option 3 has two variations. One is bit exact model to be transferred and the other is the case compilation is allowed. The second case is more UE implementation friendly for the inference, but it does not ensure the same output.*  *Observation 3: MI-Option 1 and 2 allows multiple of physical AI/ML models.*  *Observation 4: MI-Option 1 and 2 support the training and inference are both UE side. NW side training and UE side inference is not supported.*  *Observation 5: MI-Option 3 support NW side training and UE side inference. Both training and inference are UE side is not supported.*  *Observation 6: To ensure the consistency of condition between inference and training is always up to NW side responsibility.*  *Observation 7: To ensure the consistency of UE side additional condition between inference and training is UE side responsibility in MI-Option 1. Standardization may be required in MI-Option 2 and 3.*  *Observation 8: To ensure the consistency of NW side additional condition between inference and training may require explicit NW side additional condition in MI-Option 1. Standardization may be required in MI-Option 2. NW side responsibility in MI-Option 3.*  *Observation 9: One side model use case as Beam management, Positioning and CSI prediction can be same characteristics.*  *Observation 10: Two sides model use case as CSI compression requires separate discussion, although some of operation are similar to one sided model.*  *Observation 11: In MI-Option 1, model ID can some ID used for "data collection related configuration(s) and/or indication(s)"*  *Observation 12: In MI-Option 1, NW additional condition need to be informed to UE.*  *Observation 13: In MI-Option 1, UE need to inform the intention to train the models in order NW provide consistent behaviour during the training.*  *Observation 14: MI-Option 1 can be used for the following use cases.*  *- One UE side model*  *- Two-sided model of joint training at UE and NW side simultaneously*  *- Two-sided model of separate training at UE and NW side* |
| NEC[16] | *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: In the model identification procedure with data collection related configuration(s) and/or indication(s), support at least one-to-one association between model ID to report (sub)configuration ID. The information can be from NW to UE (Type B2), or from UE to NW (Type B1) with later confirmation from NW.*  *Proposal 4: The association among data collection configurations for different LCM stages needs to be provided to UE.*  *Proposal 5: In the model identification procedure, support the translation from a global model ID to a local model ID.* |
| Fujitsu[17] | *Observation-1: MI-Option 4 is dedicated to CSI compression, and its further study can depend on the progress of inter-vendor collaboration’s study in the CSI compression part.*  *Observation-2: For MI-Option 5, it is capable of identifying applicable model(s) under a certain NW-side additional condition. Model ID(s) is assigned to the selected model(s) by model monitoring for the model’s future usages.*  *Proposal-1: The procedures of model identification via model monitoring are clarified as:*  *• Applicable model(s) is selected via model monitoring under a certain NW-side additional condition*  *• NW assigns model ID(s) to the applicable model(s)*  *• The linkage between the model ID(s) and the NW-side additional conditions is setup for the model future usages*  *Proposal-2: MI-Option 5 is suggested to be studied with the three agreed MI options together for all the use cases.*  *Proposal-3: Whether to support model identification or not should be concluded first for beam management and positioning to progress its normative study.*  *Observation-3: For beam management and positioning, dataset transfer is not studied and agreed in Rel-18 SI.*  *Obeservation-4: For beam management and positioning, model training at NW side is not agreed in Rel-18 SI. Thus, there is no need to study model transfer from NW to UE in Rel-19 WI.*  *Proposal-4: For beam management and positioning, study on ML-Opiton2, ML-Option3 and ML-Option4 is suggested to be deprioritized.*  *Proposal-5: For beam management and positioning, model identification is suggested to be supported to overcome the difficulties in defining the assistance information explicitly for NW-side additional conditions.*  *Proposal-6: For beam management and positioning, ML-Option 1 and ML-Option 5 are suggested to be further studied.* |
| Continental Automotive[19] | *Proposal 1: Model ID can be applied as basis for further discussion about model identification related issues including MI-Option scenarios.*  *Proposal 2: MI-Option 5 can be one of major options for further study along with other listed three options.*  *Proposal 3: All listed MI-Options need to be kept for further study before prioritizing them.*  *Proposal 4: Mapping relation information about dataset and model with index/ID-based signaling can be applied.* |
| IDC[20] | *Observation 1: Model identification may not be necessary for collaboration Level: x, since AIML models are implementation-based and transparent to the specification.*  *Observation 2: 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 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: For MI-Option 1, the data collection configuration(s) with additional indication(s) appear to be sufficient to associate with a UE-side model, e.g., for UE-side model-level LCM.*  *Proposal 1: Model ID is not needed for one-sided UE-side model, at least for MI-Option 1.*  *Proposal 2: For AIML positioning purpose, support MI-Option 1 for model identification type B.* |
| Apple[22] | *Observation 1: MI-option 1 of model identification B can be the same procedure to handle any cell specific, site specific, configuration specific and dataset specific model.*  *Observation 2: In MI-option 1, model ID is a special case of data collection related configuration(s)/indication(s) which abstract NW side additional condition that UE is not aware of.*  *Observation 3: In MI-option 1, information transmitted from NW to UE include the configurations/indication in the RRC configurations for data collection, and for inference.*  *Observation 4: In MI-option 1, information transmitted from UE to NW include the UE’s confirmation whether UE side model supports this configurations/indication in the RRC reconfiguration complete message.*  *Proposal 1: It is up to RAN2 to define the “proactive” and “reactive” UE reporting to align the applicability condition between UE and NW. The same procedure can be used as model identification type B1/B2.* |
| Lenovo[23] | *Proposal 1: To facilitate discussion on the necessity, study the detailed information to be shared per use case and the benefit of model-ID-based LCM with the shared information.*  *Observation 1: An AI/ML model can be identified with a) information on model structure and information on dataset; b) information on model structure and weights; c) a compile binary file for deployment.*  *Proposal 2: Study the information on the data collection related configuration and dataset indication for model identification per use case.*  *Proposal 3: If a dataset for model training can be identified by a well-defined configuration/scenario, the dataset indication, e.g., a dataset ID, can be with the model identification to guarantee the consistency between training and inference.*  *Proposal 4: A dataset can be associated with a set of data collection configuration parameters, which represent the set of conditions/additional conditions of the UE, of the gNB, and even of other nodes in the network affecting the measured data.*  *Proposal 5: A set of data collection configuration parameters may include indication parameter(s) implicitly/abstractly representing a status/parameter/characteristic (that cannot be explicitly disclosed) of a UE or a gNB for data collection.*  *Proposal 6: Support procedures/signaling enabling UE and NW to exchange information related to data collection configuration parameters to associate the collected samples/group of samples with that configuration.*  *Proposal 7: Support model identification based on association of the model with the data collection configuration parameters of the samples used during the training.*  *Proposal 8: During inference, models can be selected/activated/used based on the current data collection configuration parameters and the metadata of the models identified during the identification process.*  *Observation 2: For Type A model identification, the model-related information can be shared between UE and NW based on a model ID without impact on air interface.*  *Observation 3: For Type B1 model identification, the model-related information needs to be shared between UE and NW over the air interface based on a model ID and/or model metadata.*  *Proposal 9: A local model ID and/or model metadata can be assigned and used to facilitate the model-related information sharing between NW and UE, e.g., associated scenario/configuration to collect samples for a dataset and model-ID-based LCM.*  *Observation 4: For Type B2 model identification, the model-related information can be shared between UE and NW together with model transfer over the air interface.*  *Proposal 10: For Type B2 model identification, a model ID and/or model metadata can be assigned by NW on each model transferred to UE for Model-ID-based LCM.* |
| Nokia[24] | *Observation 1: RAN1 has not concluded any details on the feasibility of offline model identification in Rel-18, but some signalling impacts (such as reporting model-IDs in UE capability report and configuring model-ID in functionalities) were discussed.*  *Proposal 1: Rel-19 study/work on model identification shall be more focused on online model identification, where the identification happens with over-the-air signaling support. Signaling designs for offline model identification (which are based on the assumptions of having inter-vendor collaborations) are not supported in the Rel-19 WI.*  *Observation 2: For MI-Option 1, for the case of online model identification based on measurement configuration(s) or data collection configuration(s), NW identifiers (which implicitly represent NW additional conditions) of the measurement configuration(s) or data collection-related configuration(s) can be reported as part of model ID to the NW.*  *Proposal 2: RAN1 to consider model identification MI-Option 1 with following high-level steps:*  *• MI-Option 1a: Model identification via NW-identifiers in measurement configurations (or via data collection configurations).*  *o The NW provides measurement configuration(s) or data collection-related configuration(s) that can be used for model identification.*  * measurement configuration(s) or data collection-related configuration(s) can be associated with NW-identifiers and global cell identities (GCIs).*  *o The UE may use the measurements corresponding to the above configurations to identify the need for any new models by assuming model training/assessment/monitoring at the UE side.*  *o The UE reports a model-ID to identify a new model together with the associated NW-identifiers (optionally GCIs) associated with the above configurations.*  *o The NW and UE can refer to the model-ID for later stages of LCM signaling.*  *o Measurement configuration can be CSI resource configuration for BM use cases or PRS resource configuration for Positioning use cases.*  *Observation 3: For MI-Option 1b, for the case of online model identification based on indications from the NW, by fixing NW-sided assumptions in certain time durations, the UE gets more opportunity to evaluate background ML model performances to identify suitable ML modes (for the observed NW assumptions). Also, as the respective background additional conditions are known at the NW and UE, the time durations can also be referred to in the signaling to identify ML models.*  *Observation 4: For MI-Option 1b, for the case of online model identification based on indications from the NW, assuming a model-ID with associated time duration(s) and other details like associated Cell(s)/PCI(s)/TRP(s)/Area information, allows NW to have a further understanding of ML model-related additional conditions.*  *Observation 5: For MI-Option 1c, for the case of online model identification based on network registration, the UE can send a list of logical ML model IDs and associated model information to the network (specific to a use case) during the network registration phase. Logical ML model information may carry similar information/indicators as in MI-Option 1a and 1b. The NW is able to validate the reported list of logical model IDs and send confirmation on the validated ML models to the UE.*  *Proposal 3: RAN1 to study other variants of MI-Option1 with following high-level steps:*  *• MI-Option 1b: Model identification referring to time duration(s)/timestamp(s) and cells/TRPs/area-related information.*  *o The NW provides time-duration(s), and other associated information such as cell(s)/TRP(s)/Area info(s) that can be used for model identification.*  *o The UE identifies any new models by performing model assessment/monitoring for the provided time duration(s) and associated cell(s)/TRP(s)/Area Info(s).*  *o The UE reports a model-ID to identify a new model and reports associated time-duration(s)/Cell(s)/TRP(s)/Area Info(s).*  *o The NW and UE can refer to the model-ID for later stages of LCM signalling.*  *• MI-Option 1c: Model identification during the network registration*  *o The UE sends a list of logical ML model IDs and associated model information to the network (specific to a sub-use case, and/or Functionality) during the network registration phase. Logical ML model information may carry similar information/indicators as in Options 1a and 1b.*  *o The NW validates the reported list of model IDs and sends confirmation on the validated ML models to the UE.*  *o In the capability signaling, the UE reports model IDs according to the validated list of ML models.*  *o The NW and UE can refer to the model-ID for later stages of LCM signalling.*  *Observation 6: For UE-sided models, model identification with the dataset transfer does not seem to be solving NW-sided additional conditions related concerns. Therefore, it may not be a solution for one-sided use cases.*  *Observation 7: For two-sided 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.*  *Observation 8: For MI-Option 3, if UE models get trained at the NW (subjected to further discussions), the NW can assign a model ID to identify the model associated with the model transfer during the model transfer process and that model ID can be used later in the LCM.*  *Observation 9: For MI-Option 3, as model transfer discussions happen separately and not only in RAN1, it is hard to conclude yet at which stage model transfer related model identification shall, and if, occur.* |
| ETRI[25] | *Proposal 1: Model-ID-based LCM can be integrated with Functionality-based LCM by using model ID for LCM operations.*  *Observation 1: Utilizing functionality-based LCM with data collection-related configurations and/or indications results in the generated dataset becoming dependent on each functionality.*  *Observation 2: The model identification process requires a detailed discussion of specific procedures, and it may reveal proprietary information about the AI/ML models.*  *Proposal 2: In the case of MI-Option 1, after configurations or indications, data samples are stored, and a dataset is created with a dataset ID.*  *Proposal 3: In the case of MI-Option 2, the NW can provide dataset information corresponding to the configured functionality to the UE as an additional condition.*  *Observation 3: Delivering datasets within defined functionalities allows the avoidance of proprietary concerns.* |
| AT&T[26] | *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: Model identification provides model-level management by NW of UE-side and UE-part of two-sided models, which may provide benefits at least in the following scenarios:*  *• UE side models with model transfer*  *• Pairing of two-sided models*  *• To ensure consistency between training and inference regarding NW-side additional conditions*  *• To provide enhanced performance on certain scenario/configuration/datasets.*  *• To enable more granular (model-level) performance monitoring at NW*  *• Target performance of the model may be provided to NW during/after model identification.*  *• To provide awareness at NW on UE-side model switching interruption.*  *Proposal 4: 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 5: Regarding MI-Option 2 (Model identification with dataset transfer) of model identification type B, RAN1 further study the following aspects:*  *• Relationship between model ID and dataset*  *• 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 2*  *Note: whether MI-Option 2 is needed or not is a separate discussion.*  *Proposal 6: Regarding MI-Option 3 (Model identification in model transfer from NW to UE) of model identification type B, RAN1 further study the following aspects:*  *• 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 3*  *Note: whether MI-Option 3 is needed or not is a separate discussion.*  *Proposal 7: Regarding MI-Option 5 (Model identification via model monitoring) of model identification type B, RAN1 further study the following aspects:*  *• Relationship between model ID and model monitoring configurations*  *• 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 5*  *Note: whether MI-Option 5 is needed or not is a separate discussion.* |
| Qualcomm[27] | *Proposal 1: To facilitate future discussions and for the purpose of brevity, refer to an identifier associated with “data collection related configuration(s) and/or indication(s)” as dataset ID.*  *Proposal 2: For model identification with data collection related configuration(s) and/or indication(s) (MI-Option 1), identify the following options for model identification Type B:*  *• NW-side-initiated model identification*  *• UE-side-initiated model identification*  *Observation 1: For Model identification with data collection related configuration(s) and/or indication(s), e.g., dataset ID, (MI-Option 1), and for UE-side-initiated model identification, different UE vendors and/or different UE types may use different model IDs for the same set of NW-side additional conditions, i.e., for the same dataset ID.*  *Proposal 3: For Model identification with data collection related configuration(s) and/or indication(s), e.g., dataset ID, (MI-Option 1), and for NW-side-initiated model identification Type B, dataset ID may be regarded as a (logical) model ID, and the NW’s signaling of a dataset ID may be regarded as a trivial “model identification”. In other words, there’s a one-to-one relationship between model ID and dataset ID.*  *Proposal 4: To facilitate the discussion, RAN1 studies the following options as starting point for model identification type A 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 exchange*  *• MI-Option 3: Model identification with model exchange between NW-side and UE-side*  *• MI-Option 4: Model identification via standardization of reference models. (for CSI compression)*  *• MI-Option 5: Model identification via standardized dataset*  *• FFS: The boundary of the options*  *• Note: the names (MI-Opton1, MI-Option 2, MI-Option 3, MI-Option 4, MI-Option 5) are used only for discussion purpose*  *Note: other options are not precluded*  *Proposal 5: For model identification with data collection related configuration(s) and/or indication(s) (MI-Option 1), identify the following options for model identification Type A:*  *• NW-side-initiated model identification*  *• UE-side-initiated model identification*  *Proposal 6: For Model identification with data collection related configuration(s) and/or indication(s), e.g., dataset ID, (MI-Option 1), and for NW-side-initiated model identification Type A, dataset ID may be regarded as a (logical) model ID, and the NW’s signaling of a dataset ID may be regarded as a trivial “model identification”. In other words, there’s a one-to-one relationship between model ID and dataset ID.*  *Observation 2: Model identification may be used for the following purposes:*  *- To achieve alignment on the NW-side additional condition between NW-side and UE-side*  *- To identify models during and after model transfer*  *- Along with data collection related configuration(s) and/or indication(s) and/or dataset transfer*  *- For pairing of two-sided models*  *Observation 3: Models developed and trained for a particular scenario/configuration/dataset may provide better performance than a generalized model that covers a wide range of scenarios/configurations/datasets.*  *Proposal 7: Target performance may be associated with a model and aligned between the NW side and UE side during model identification.*  *Proposal 8: Model identification provides means for (1) identifying a stronger performance during model identification for a scenario/configuration/dataset, (2) correspondingly set a stronger performance target for the scenario/configuration/dataset, (3) and monitoring for the stronger performance via model monitoring during inference.*  *Proposal 9: Model identification may be used for selected scenario/configuration/datasets for enhanced performance within functionality-based LCM.*  *Observation 4: Model identification requires minimal-to-none inter-vendor collaboration effort and does not have a scalability issue.*  *Observation 5: Additional spec impact for model identification is minimal.*  *Proposal 10: Support model identification in Rel-19 normative work.* |
|  | *Proposal 1: 3GPP should consider the framework to support scenario/site specific model.*  *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.*  *Proposal 2: Proponent of MI-Option5 should clarify the difference between MI-Option1 and MI-Option5.*  *Observation 2: Functionality identification and model identification can be differentiated according to whether new AI/ML related ID is introduced or not.*  *Observation 3: The following approaches are considered to prepare the model specific to NW side additional condition(s)*  * UE side training*  *・ Via offline coordination*  *・ Via information/indication associated with additional conditions from NW to UE*  *- Via explicit information on NW side additional condition*  *- Via ID indication (e.g., ID representing data collection configuration associated with specific additional conditions)*  * NW side training*  *・ Via model transfer*  *Observation 4: Consistency assisted by monitoring can be categorized into the following two types:*  * UE autonomous monitoring. UE constantly monitor the performance to check the consistency.*  * NW initiated monitoring. NW initiates UE to monitor the performance when NW side additional condition is changed.*  *Observation 5: The following approaches are considered to check consistency between NW side additional conditions and UE side model for inference:*  * Via information/indication associated with additional condition from NW to UE*  *・ Via explicit information on NW side additional condition*  *・ Via ID indication (e.g., model ID, dataset ID, data collection configuration ID)*  * Via monitoring of model/functionality*  *・ Via UE autonomous monitoring*  *・ Via NW initiated monitoring*  *Observation 6: After model identification, consistency between NW side additional condition and UE side model for inference can be aligned via model ID indication from NW.*  Proposal 3: Conclude the pros and cons of model identification type as Table 1.  Table 1. Characteristic of model identification types.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Aspects＼Type** | **Type A** | **MI-Option1** | **MI-Option2** | **MI-Option3** | **MI-Option4** | **MI-Option5** | | How to prepare the model specific to NW side additional condition(s) | Offline coordination | Data collection via dataset collection configuration | Model transfer | Dataset transfer | Inapplicable | Inapplicable | | How to ensure consistency between NW side additional conditions and UE side model | ID indication | ID indication | ID indication | ID indication | ID indication | ID indication | | Two-sided model use case applicability | Applicable | Inapplicable | Applicable | Applicable | Applicable | Inapplicable | | Challenges | Requires offline coordination | - | Requires model transfer | Requires dataset transfer | - | - |   Observation 7: if the following two points are not desirable, model identification should be considered to handle NW side additional conditions.   * Proprietary issue due to providing information of NW side additional conditions to UE * Not support of NW additional condition specific model preparation   Proposal 4: Conclude the pros and cons of approaches without model identification as Table 2.  Table 2. Characteristic of approaches handling NW side additional condition without model identification.   |  |  |  |  | | --- | --- | --- | --- | | **Aspects＼Type** | Information and/or indication on NW side additional conditions is provided to UE | Consistency assisted by monitoring | | | UE autonomous monitoring | NW initiated monitoring | | How to prepare the model specific to NW side additional condition(s) | Explicit NW additional condition information from NW to UE | Infeasible | Infeasible | | How to check consistency between NW side additional condition and UE side model | Explicit NW additional condition information from NW to UE | UE autonomous monitoring | NW initiated monitoring | | Challenges | Proprietary issue | Requires constant monitoring at UE side | Monitoring is required when gNB deployment is changed | |
| DCM[28] |
|  |  |
|  |  |

#### **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.

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

In general, companies still have quite divergent views on whether model identification should be supported or not in Rel-19.

#### **Proposal 2.1.1**

Many tdocs continue discussing the relationship of functionality-based LCM and model-ID-based LCM, e.g., whether a unified LCM consisting of both of them, whether model-ID-based LCM is working within functionality-based LCM.

In the last meeting, the group got a relatively stable proposal for this topic. Meanwhile, moderator feels that such kind of clarification can facilitate further discussion.

Thus, moderator suggests to continue discussing this issue based on this version in Proposal 2.1.1.

**Proposal 2.1.1**

**For Rel-19, model-ID-based LCM (if supported) refers to using model ID for LCM operations within a functionality.**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | Based on the discussion in the last meeting, at least 19 companies seem ok with the proposal:   * CATT, Huawei, Lenovo, Xiaomi, Fujitsu, ETRI, QC, Ericsson, Panasonic, Samsung, Spreadtrum, CMCC, OPPO, NEC, CEWiT, Intel, CICTCI, New H3C, MTK,   @ZTE: In the last meeting, “or across functionalities” was suggested by ZTE. From moderator’s perspective, Proposal 2.1.1 is only talking about “within a functionality”, not touching how the model ID is used “across functionalities” that can be discussed separately. The current version is not conflicting with your proposal and should be ok. |
| Panasonic | Basically Support. Not required to say "for Rel-19" as it is more generic. I also support ZTE comment that a model can be applied across multiple of functionalities. Then separate bullet discussion can be added like following.  - ~~For Rel-19, m~~Model-ID-based LCM (if supported) refers to using model ID for LCM operations within a functionality.  - A model can be applied among LCM operations over multiple functionalities. |
| Samsung | Thank you FL. We are agree with the proposal. |
| Mavenir | We support this proposal |
| NTT DOCOMO | Fine with the proposal. Another question is whether model ID based LCM could refer to using dataset ID for LCM operations within a functionality. In our view, if newly AI/ML related ID (e.g., model ID, dataset ID) is introduced, LCM can be categorized into model ID based LCM. If the existing ID (e.g., CSI-ResourceConfigId) is used or no ID is used, LCM can be categorized into functionality based LCM. |
| New H3C | OK |
| ZTE | Thanks moderator for the great effort. We think it may be better to first discuss the model identification options, and the come back to this proposal since currently it is clear what the model ID refers to yet. For example, depending on different model identification options, the so-called model ID may be referred to Model ID, dataset ID or consistency ID.  Regarding whether the model ID can be applied across Functionalities, we think it depends on several aspects, e.g., the granularity of Functionality design, definition of model ID, etc. For example,   * If the model ID is dataset ID, only dataset may be applied for several functionalities. One example is the dataset for beam case 2 can also be applied for beam case 1; another example is the dataset for CSI prediction can also be applied for CSI compression. * If the model ID is additional condition ID, it may also be applied for several Functionalities. For example, beam shape mentioned by several companies may be the additional condition for both beam case 1 and case 2.   Overall, we prefer to first discuss the model identification options and come back to this. But if majority companies prefer to agree the proposal, we suggest the following updates.  ***For Rel-19, model-ID-based LCM (if supported) refers to using model ID for LCM operations ~~within a functionality~~.*** |
| Huawei, HiSilicon | Same feeling as ZTE. Our thinking is that “within a functionality” is not relevant to model ID based LCM. Model ID based LCM may or may not introduce the concept of “functionality”  **For Rel-19, model-ID-based LCM (if supported) refers to using model ID for LCM operations ~~within a functionality~~** |
| Xiaomi | Generally, we are OK with ZTE’s proposed way forward.  **For Rel-19, model-ID-based LCM (if supported) refers to using model ID for LCM operations ~~within a functionality~~**  At current stage, we could keep the door for “within a functionality” or “across functionalities” open. When there is sufficient progress in the model identification options and functionality identification, we could step further. |
| Fujitsu | We are fine with the proposal in principle. Regarding model ID, we think it would be better to have further clarification on:  Alt-1: the model-ID is global ID only  Alt-2: the mode-ID is local ID only  Alt-3: the model-ID can be either global ID or local ID  With this clarification, the relationship between model-ID and dataset-ID or additional-ID can be concluded easily. |
| vivo | We prefer to discuss Proposal 2.1.2 first before we touch detailed signaling of model identification. |
| QC | OK |
| mtk | We still feel that it would be better to have an unified LCM procedure, not to distinguish it is function based or model ID based. Even though function based LCM, any ID is needed |
| Nokia | OK with the direction. However, it is good to discuss a bit on details of model identification before this agreement. |
| NEC | We are OK with the proposal. |

#### **Proposal 2.1.2**

Based the tdocs, there are different understanding on the model identification for MI-Option 1. It is critical to have a common understanding so that the group can continue discussion for the necessity/requirement of model identification based on a common basis. Otherwise, different companies may use the same terminology “model identification” to refer to different things. Thus, Proposal 2.1.2 is suggested for further discussion

* These are just simplified examples. Other variants are not listed and many details are omitted since they don’t have much impact on this clarification.

**Proposal 2.1.2**

**From RAN1 perspective, Mechanism#1 with the following exemplary procedures is [one example of model identification MI-Option 1] [NOT model identification]**

* **NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s)**
* **UE collects the corresponding data based on which one or more AI model(s) are trained**
* **UE reports to NW that the AI model(s) is applicable to these associated ID(s)**
* **NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | * Some companies think the above mechanism#1 is one kind of model identification since the associated ID can be regarded as an implicit model ID. * In contrast, some other companies think mechanism#1 is not model identification (just a method to ensure the consistency) since no model ID is explicitly assigned/defined.   Thus, it would be beneficial to achieve a common understanding within the group before we can achieve on any consensus on the necessity of MI-Option 1.  It should be the common understanding within the group that Mechanism#2 with the following exemplary procedures is one example of model identification MI-Option1:   * NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s) * UE collects the corresponding data based on which one or more AI model(s) are trained * UE reports to NW that the AI model(s) is applicable to these associated ID(s). And then NW assigns model ID(s) to the AI model(s) * NW indicates the LCM operation and its associated model ID. And then UE selects the suitable AI model for the corresponding operation. |
| Panasonic | - In both Mechanism#1 and Mechanism#2, the last bullet " •NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s)." is not the part of model identification but rather within " model selection, activation, deactivation, switching, and fallback operation."  - The difference between Mechanism#1 and Mechanism#2 is associated ID(s) is assigned at the time of data collection or assigned at the time after training. Even in Mechanism#2, model ID(s) is related to associated ID(s) in the data collection. Then renaming of the associated ID(s) to model ID(s) can be allowed or not. With generic use case level discussion, this difference is not required to be debated and can be combined with the single procedure. Therefore, my thinking is following.  From RAN1 perspective, Mechanism#1 with the following exemplary procedures is ~~[~~one example of model identification MI-Option 1~~]~~ ~~[NOT model identification]~~   * NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s) * UE collects the corresponding data based on which one or more AI model(s) are trained * UE reports to NW that the AI model(s) is applicable to these associated ID(s). It is FFS whether NW assigns model ID(s) to the AI model(s) specifically. * ~~NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).~~ |
| Samsung | We share the concern. There are two things combined here. First issue is to ensure consistency between model training and inference on NW-side additional conditions. Second issue is for model-level LCM assistance (management) by the network.  **From RAN1 perspective, Mechanism#1 with the following exemplary procedures is [one example of model identification MI-Option 1] ~~[NOT model identification]~~**  **Step1: For data collection, NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s)**  **Step2: UE collects the corresponding data based on which one or more AI model(s) are trained**  **Step 3: For inference, NW signals the configuration/indications with an associated ID(s). ~~And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).~~**  **Step 4: UE ~~reports to~~ identifies to the NW ~~that~~ the applicable AI model(s) ~~is applicable~~ to the~~se~~ associated ID(s)**  **Note: Step 1, Step 2, and Step 3 are to ensure consistency between model training and inference regarding NW-side additional conditions. Step 1, Step 2, and Step 3 are applicable to functionality-based LCM. Step 4 is additional step for UE’s model identification (model identification Type B1) for model-ID-based LCM.** |
| NTT DOCOMO | Current description includes both model identification and functionality identification. If the associated ID is new AI/ML related ID (model ID or dataset ID), Mechanism#1 and Mechanism#2 can be called model identification. If associated ID is the existing ID (e.g., CSI-ResourceConfigId), Mechanism#1 is functionality identification. To clarify that, we suggest the following note in the proposal.  Note: if the associated ID is the existing ID (e.g., CSI-ResourceConfigId), the above procedure is called functionality identification. |
| New H3C | OK with samsung’s modification |
| ZTE | First of all, we agree with other companies that MI-Option1 is not a model identification procedure, instead it is a mechanism for additional condition.  Secondly, based on our understanding, there are at least two variations of the MI-Option1 as shown below. Example#1 is mainly for model-ID based LCM and network indicates model ID to manage UE’s model. Example#2 is mainly for Functionality-based LCM, where network manages the Functionality and UE translates Functionality management to model management based on the additional condition indication.  Overall, we propose the following updates.  **From RAN1 perspective, Mechanism#1 with the following exemplary procedures is ~~[one example of model identification MI-Option 1] [~~NOT model identification~~]~~**  **Example#1**   * **StepA1: NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s) during model training phase** * **StepA2: UE collects the corresponding data, based on which one or more AI model(s) are trained** * **StepA3: UE reports to NW that the AI model(s) is ~~applicable~~ associated to these associated ID(s)**   **Example#2**   * **StepB1: NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s) during model training phase** * **StepB2: UE collects the corresponding data, based on which one or more AI model(s) are trained** * **StepB3: NW signals the configuration/indications with an associated ID(s) during inference phase. UE reports to NW that the supported Functionalities. And then UE accordingly selects a suitable AI model for the following AI/ML Functionality operation(s).** |
| Lenovo | The procedure is okay assuming that we are discussing UE-sided model which is trained at the UE side. For a more general cases, the procedure needs to be modified a bit. So, for now we suggest:  **Proposal 2.1.2**  **From RAN1 perspective, Mechanism#1 with the following exemplary procedures is [one example of model identification MI-Option 1 for UE-sided models trained at the UE side] [NOT model identification]**  Furthermore, we believe the above procedure **is** one kind of model identification. We note that: **Model identification:** A process/method of identifying an AI/ML model for the common understanding between the NW and the UE  So, any procedure that can help NW and the UE to have the same understanding on how to train the model with the configured data collection can be considered as a model identification process **~~even~~** ~~if there is not Model-ID~~.  Also, in another perspective, the associated IDs (of data collection) that are assigned to the model (In Mechanism#1) can be considered as the ID of the model. ~~So in a way we have Model-ID as well~~. |
| Huawei, HiSilicon | **Firstly**, we think the key motivation is to enable UE to categorize NW side additional conditions. Whether it is subject to model identification procedure is a second step. In our understanding, the associated data categorization ID should be local rather than global, while the model ID is globally unique as agreed, so Mechanism#1 is NOT model identification. To avoid tangling this controversial point, we may first discuss the step of Mechanism#1 before classifying it. => change on main bullet  **Secondly**, as mentioned in the first point, we may take local ID as a starting point. => add a 5th sub-bullet  **Thirdly**, as addressed in our Tdoc R1-2402027, the NW may probably have no knowledge on how to perform data categorization – assuming a gNB may dynamically change the vertical beam angle among the range 5°-15°, how to categorize these continuous values into discrete data IDs? If it takes the 0.1° as the granularity, it may distinguish them into 100 IDs. When additionally consider other factors such as beam width, beam mapping order, etc. the overall number of IDs can be tremendous, which we believe is helpless to UE to train the model. => add a 6th sub-bullet  **Fourthly**, for the “UE report to NW that AI models is applicable”, we are not confident that this procedure is needed. NW can indicate the same data categorization ID for training and for inference, so UE can schedule the same trained model for inference. => move 3rd sub-bullet to FFS  **From RAN1 perspective, study Mechanism#1 with the following exemplary procedures ~~is [one example of model identification MI-Option 1] [NOT model identification]~~**   * **NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s)** * **UE collects the corresponding data based on which one or more AI model(s) are trained** * **FFS UE reports to NW that the AI model(s) is applicable to these associated ID(s)** * **NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).** * **Regarding associated ID(s), take cell specific ID as a starting point.** * **Address the issue of mismatch between NW side and UE side on the categorization rule to map additional conditions into separate IDs.** |
| Xiaomi | * Firstly, we don’t think Mechanism#1 is model identification. It is kind of utilizing assistance information to keep consistency between training and inference. It can be applied to the functionality identification case and model identification case. And most of all, there is no model ID determination step in this machanism. * For Mechanism #2，we think it is kind of model identification example except for the last bullet. The last bullet illustrates how to select the model. We think may be it is a separate issue with model identification. At current stage, the model selection can be performed with assistance information from NW or directly performed by the NW. |
| Fujitsu | If the model-ID is assumed as global ID only, and “the associated ID(s) in data collection” is a global ID as well, the above mechanism#1 is a kind of model identification. Here, the associated ID can be defined with either Alt-1 or Alt-2:  Alt-1: the associated ID is defined as a local ID associated with a global ID (e.g. GCI) during data collection.  Alt-2: the associated ID is defined as a global ID directly during data collection.  It is possible that a model could be applicable for multiple datasets (with different associated IDs), but it can be regard as a model-ID-based mechanism as well. In this case, multiple associated IDs = one model-ID. |
| vivo | We believe this procedure is not Model identification procedure. As we illustrated in our paper, dataset categorization (or data categorization ID) and model ID have different underlying logic.   * Dataset categorization 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 dataset categorization.   Directly using data categorization information as model ID is not future-proof for cases where real model-level awareness is needed.  **From RAN1 perspective, Mechanism#1 with the following exemplary procedures is ~~[one example of model identification MI-Option 1] [~~NOT model identification~~]~~**   * **NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s)** * **UE collects the corresponding data based on which one or more AI model(s) are trained** * **UE reports to NW that the AI model(s) is applicable to these associated ID(s)** * **NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).** |
| QC | The main intention of MI-Option 1 is to help clarify **the link and association between model ID and data collection related configuration(s) and/or indication(s)**, which is not clear from Proposal 2.1.2 in its current form. In our view there can be two flavors with regards to the connection between model ID and data collection related configuration(s) and/or indication(s), based on whether we are considering NW-side initiated model identification or UE-side initiated model identification.  For NW-side-initiated model identification Type B, the identifier associated with data collection related configuration(s) and/or indication(s) may be regarded as a (logical) model ID, and the NW’s signaling of a dataset ID may be regarded as a trivial “model identification”. In other words, there’s a one-to-one relationship between model ID and dataset ID. On the other hand, for UE-side-initiated model identification, different UE vendors and/or different UE types may use different model IDs for the same set of NW-side additional conditions, i.e., for the same identifier associated with data collection related configuration(s) and/or indication(s).  So, we suggest the following update:  **Updated Proposal 2.1.2**  **From RAN1 perspective, Mechanism#1 with the following exemplary procedures is [one example of model identification MI-Option 1] ~~[NOT model identification]~~**   * **NW signals the data collection related configuration(s) and/or indication(s), ~~and the~~ in the form of associated ID(s)** * **Each associated ID with the data collection related configuration(s) and/or indication(s) can be regarded as a (logical) model ID assigned by NW.** * **UE collects the corresponding data for each associated ID based on which ~~one or more~~ specific AI model(s) are trained for each associated ID.** * **~~UE reports to NW that the AI model(s) is applicable to these associated ID(s)~~** * **During UE capability, UE indicated the supported model IDs to the NW**   **NW signals the ~~configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).~~ model ID that the UE should use during inference, at least based on UE capability.** |
| Nokia | Ok with the high-level direction of the proposal, but some modifications are needed to make things accurate.  **From RAN1 perspective, ~~Mechanism#1 with~~ the following exemplary procedures can be supported for ~~is [one example of~~ model identification MI-Option 1~~] [NOT model identification]~~**   * **NW signals the data collection ~~related~~ configuration(s) ~~and/or indication(s)~~, and indicates ~~the~~ associated ID(s) of data collection configuration(s).**    + **Note: the associated ID(s) may serve the purpose of data set categorization with respect to NW-additional conditions** * **UE collects the corresponding data based on which one or more AI/ML model(s) are trained** * **UE reports to NW that the AI/ML model(s), via assigning and reporting a model IDs, and each model-ID is related to ~~applicable to~~ these associated ID(s)** * **~~NW signals the configuration/indications with an associated ID(s). And then UE accordingly selects a suitable AI model for the following AI/ML operation(s).~~** |
| NEC | We think the procedure in the proposal may not be model identification procedure if the associated ID is not the model ID. |

#### **Proposal 2.1.3**

Based on the tdocs, many companies suggest to further study MI-Option 4. Regarding MI-Option 4, some tdocs say it doesn’t need any mechanism for model identification as the reference models and the associated IDs are pre-defined. Meanwhile, according to some other tdocs, UE needs to report which reference model(s) is supported, which also belongs to model identification. Thus, Proposal 2.1.3 is suggested to clarify MI-Option 4 itself and whether some potential spec impact is needed or not.

**Proposal 2.1.3**

* **For MI-Option 4 (if supported), the structure and parameters of reference AI model(s) and the associated model ID(s) (if needed) are standardized in 3GPP**
* **For MI-Option 4 (if supported), there is no additional spec impact other than the following aspect from RAN1 perspective**
  + **UE reports to network which reference AI model(s) is supported if multiple reference AI models are standardized**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | If only model structure is standardized in 3GPP, then MI-Option 4 is a special case of model transfer (i.e., case z4 where the known structure is assumed). |
| Panasonic | If both structure and parameters of reference model(s) are standardized, I agree it is one of the model identifications.  We support the proposal. |
| Samsung | We can first agree on the first bullet. For the second bullet, it is too early to conclude that. As an example, a UE may identify its model (Type B1) by indicating the compatible reference model(s). |
| Mavenir | We agree with the first bullet. |
| NTT DOCOMO | Standardizing reference model structure and parameters for actual deployment could be RAN1 works instead of RAN4 works, if MI-Option 4 is supported. From that perspective, it is safe to avoid concluding that no RAN1 spec impacts is expected for reference model design. Hence, we suggest the following update.   * **For MI-Option 4 (if supported), there is no additional spec impact other than the following aspect from ~~RAN1~~ over the air signaling perspective**   **UE reports to network which reference AI model(s) is supported if multiple reference AI models are standardized** |
| New H3C | OK in general |
| ZTE | Ok with this proposal. |
| Lenovo | Suggest to discuss the reference model in 9.1.3.2 there. |
| Huawei, HiSilicon | Firstly, MI-Option 4 does not have a model identification/model ID assignment procedure – its model IDs are specified by spec. Thus NW and UE can skip the model identification procedure and directly refer the specified model IDs for pairing procedure in the inference phase. So, we need to revisit the procedure of model identification for MI-Option 4.  Secondly, whether there is additional procedure is unclear, so the main bullet is removed, while only the sub-bullet is kept.   * **For MI-Option 4 (if supported), the model identification procedure is revisited as “**The model may be assigned/specified with a model ID during the model identification**”** * **For MI-Option 4 (if supported), the structure and parameters of reference AI model(s) and the associated model ID(s) (if needed) are standardized in 3GPP** * **~~For MI-Option 4 (if supported), there is no additional spec impact other than the following aspect from RAN1 perspective~~**   + **For MI-Option 4 (if supported), UE reports to network which reference AI model(s) is supported if multiple reference AI models are standardized** |
| Xiaomi | It seems MI-Option 4 is related to the discussion of two-sided model. Thus, we have similar suggestion with Lenovo, it can be discussed in section 9.1.3.2 or wait for sufficient progress in section 9.1.3.2 |
| Fujitsu | We also think the first bullet can be agreed first. |
| vivo | First bullet is fine. |
| QC | This option was motivated by the CSI compression use case, and in our view, it would be more sensible to wait for more elaboration/conclusion from 9.1.3.2, before further discussions related to this option. This is based on one of the options agreed by RAN1 in 9.1.3.2 agenda item (RAN1 #116), and we can discuss this further when we have more clarity coming out of 9.1.3.2 discussions. |
| Nokia | It is not clear how specifying a structure of a model is identifying NW-additional conditions. At least for one sided model, there are missing parts that needs some clarifications for MI-Option4. For two-sided models, it would be good to discuss some details in 9.1.3.2 |
| NEC | If option 4 is supported, it is needed to specify the structure and parameters of reference AI model(s) and the associated model ID(s) as said in the first bullet. |

#### **Proposal 2.1.4**

Many tdocs are discussing the applicable use cases. Companies have similar views except for the following aspects:

**Whether MI-Option 1 can be used for two-sided model or not**: Most companies think it can be only used for one-sided model. Meanwhile, two companies think it is also workable for two-sided model. Let’s consider the three training collaboration types for CSI compression

* Training collaboration type 1:
  + If the encoder and decoder are trained at UE side, then some additional mechanism is needed to transfer/deliver the decoder to gNB to complete the model identification procedure. Thus, MI-Option 1 itself seems not sufficient for this case.
  + If the encoder and decoder are trained at NW side, then some additional mechanism is needed to transfer/deliver the encoder to UE to complete the model identification procedure. Thus, MI-Option 1 itself seems not sufficient for this case.
* Training collaboration type 2:
  + Type 2 over the air interface for model training is deprioritized in Rel-18 study.
* Training collaboration type 3
  + No matter UE first or NW first training is used, some additional mechanism is needed to transfer/delivery the training dataset for encoder or decoder. Thus, MI-Option 1 itself seems not sufficient for this case.

Based on the above discussion, moderator tends to agree with the majority companies that MI-Option 1 is applicable to one-sided model.

**Controversial views on MI-Option 5**: Let’s have a separate discussion for MI-Option 5 in other proposal.

Thus, Proposal 2.1.4 is suggested for further discussion to clarify the potential applicable use cases of different options.

**Proposal 2.1.4**

**Regarding model identification type B,**

* **MI-Option 1 can be used for UE-sided model**
* **MI-Option 2 can be used for one-sided model and two-sided model**
* **MI-Option 3 can be used for UE-sided model and two-sided model**

**Note: the necessity of model identification is a separate discussion.**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | Support |
| Samsung | Ok.   * **MI-Option 2 can be used for ~~one~~ UE-side~~d~~ model and two-sided model** |
| NTT DOCOMO | Fine with the proposal. |
| New H3C | OK in general |
| ZTE | First of all, as we discussed in proposal 2.1.2, MI-Option 1 may not be a model identification procedure. We suggest to come back to MI-Option 1 until we reached consensus on proposal 2.1.2.  If companies agree that MI-Option 1 is not a model identification procedure, we agree with moderator that MI-Option 1 can only be applied for UE-side model.  While for MI-Option 2 and MI-Option 3, the only reason to support these two options is address the model pairing of two-sided model. More justification is needed to clarify the application of MI-Option 2 and MI-Option 3 for UE-sided model. Although we agree that they can be used for one-sided model, but the main focus and main motivation to have MI-Option 2 and MI-Option 3 should be two-sided model.  Overall, we propose the following.  ***Regarding model identification type B,***   * ***~~MI-Option 1 can be used for UE-sided model~~*** * ***MI-Option 2 can be used at least for ~~one-sided model and~~ two-sided model*** * ***MI-Option 3 can be used at least for ~~UE-sided model and~~ two-sided model***   ***Note: the necessity of model identification is a separate discussion.*** |
| Lenovo | We believe **MI-Option 1** can be used for both one-side and two-sided models. More precisely, method of Option 1 associates the collected data with identifiers during data collection. The dataset constructed for training of two-sided models, then, have the association with the identifiers of data collection. The trained model, likewise, can also be associated with the identifiers of the dataset (data collection). So, although the model identification is happing based on the dataset exchange, it is still based on MI-Option1.  We also note that, for two-sided models, during data collection, the UE may also associate data with some information regarding the UE as well. The developed model later can be associated with the IDs associated with the data (by the NW or UE). Model selection/activation would be similar to the Mechanism-1 except that UE and NW exchange the required information.  The following is a high level procedure assuming that the data collection is happening at the UE side:   * **NW signals the data collection related configuration(s) and/or indication(s), and the associated ID(s)** * **UE collects the data and associates it with information received from the NW and also may add information/indications related to UE configuration(s)/states.** * **The generated dataset (with associated IDs during data collection) is used to train one or more AI model(s) (it could be any training mechanism) and both UE-parts and NW-parts are associated with the data collection ID(s)** * **During inference: NW and UE exchange configuration/indications with an associated ID(s). Then UE and NW accordingly select a suitable UE-part and NW-part for the following AI/ML operation(s).** |
| Huawei, HiSilicon | We share similar view as ZTE.  **Firstly**, as mentioned in our replies to Proposal 2.1.2, it is our understanding that MI-Option 1 does not belong to model identification, if the data categorization ID is used for UE sided model.  **Secondly**, we do not see clear necessity of MI-Option 2/3 for UE sided model.  **Regarding model identification type B,**   * **FFS MI-Option 1 ~~can be used for UE-sided model~~** * **MI-Option 2 can be used for at least ~~one-sided model and~~ two-sided model** * **MI-Option 3 can be used for at least ~~UE-sided model and~~ two-sided model**   **Note: the necessity of model identification is a separate discussion.** |
| Xiaomi | We support this direction.  There is Only a small concern on the terminology in this proposal. In our understanding “UE sided model” in the first subbullet and third subbullet may include the one sided model and UE part of two-sided model. But the intension is only for one sided model on UE side, right? We suggest the following update.  **Proposal 2.1.4**  **Regarding model identification type B,**   * **MI-Option 1 can be used for ~~UE-sided model~~ one sided model** * **MI-Option 2 can be used for one-sided model and two-sided model** * **MI-Option 3 can be used for ~~UE-sided model~~ one sided model and two-sided model**   **Note: the necessity of model identification is a separate discussion.** |
| Fujitsu | We share the similar view as that of ZTE. It is quite unclear how NW can own dataset for training UE-side model. Without UE-side additional conditions, the feasibility of NW-side training is questionable. This challenge cannot be crossed by either MI-Option 2 or MI-Option 3. |
| QC | OK |
| Nokia | We suggest to list down the details on options before this. |

#### **Proposal 2.1.5**

The proponent of MI-Option 5 clarifies the procedure as below:

* Applicable model(s) is selected via model monitoring under a certain NW-side additional condition
* NW assigns model ID(s) to the applicable model(s)
* The linkage between the model ID(s) and the NW-side additional conditions is setup for the model future usages

Meanwhile, many companies raise concerns on MI-Option 5. Thus, Proposal 2.1.5 is suggested to further study MI-Option 5 so that the group can have better understanding.

**Proposal 2.1.5:**

**Further study MI-Option 5 (including feasibility) from the following aspects as a starting point:**

* **Whether it is workable for one-sided model and/or two-sided model?**
* **Will local model ID or global model ID be assigned?**
* **Whether the models identified in one cell can be known to other cells?**
* **What’s the relationship between MI-Option 5 and the normal performance/model monitoring?**
* **What’s the relationship with other options (e.g., MI-Option 1)?**
* **UE complexity and latency especially when there is no suitable AI model for one scenario/cell**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | Support |
| Lenovo | Suggest to discuss MI-Option 1, 2 and 3 firstly. |
| Huawei, HiSilicon | We think MI-Option 5 can be considered with lower priority, before we sort out whether it belongs to model identification. |
| Fujitsu | We are fine with the proposal in general to have further study from the aspects summarized by FL.  For UE-side model, considering the difficulties of obtain sufficient NW-side additional conditions, the proposed dataset ID of MI-Option1 may be difficult to ensure a high degree of training-inference consistency. For example, for beam management, besides beam pattern of SetA/SetB, beam downtilt angle, beam width, antenna layout, polarization, antenna distance, nonlinear RF parameters, etc. may all have impact on the features of the radio data. The complexity of the radio features is the root of where AI gain coming from and is difficult to be indicated by one or several associated ID(s).  We think MI-Option5 provides a way to ensure the training-inference consistency from performance perspective. After the AI/ML performance being checked, the model ID(s) are assigned to the applicable models online.  It can be as an independent model identification mechanism or can be worked with MI-Option 1 together to mitigate the drawback of with MI-Option 1. |
| QC | Obviously, the trial-and-error approach based on monitoring is largely inefficient, as it will incur overhead in terms of latency and UE power consumption. Additionally, it is not guaranteed that an appropriate AI/ML model will eventually be found for the underlying set of NW-side additional conditions. Maybe a quick poll would be helpful to see which companies are supportive of this option, and how they address the serious feasibility and efficiency issues associated with this option. |
| Nokia | Not needed. |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.6**

In the tdocs, different purposes/usages of model identification are discussed. Different options may only achieve some of these purposes, or serve some of the usages. Thus, Proposal 2.1.6 is suggested for further discussion in order to summarize the capabilities of different options

**Proposal 2.1.6**

**Observation: Regarding the purpose/suage of different options, we have the following observations:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Purpose/usage** | **MI-Option 1** | **MI-Option 2** | **MI-Option 3** | **MI-Option 4** |
| **Can ensure consistency between training and inference (i.e., network additional alignment)** | **Y** | **Y** | **Y** | **N** |
| **Can support scenario/site/cell specific model** | **Y** | **Y** | **Y** | **N** |
| **Can facilitate two-sided model pairing** | **N** | **Y** | **Y** | **Y** |
| **Is along with Model transfer** | **N** | **N** | **Y** | **N** |
| **Can support different AI model with different capabilities (e.g., different performance, different interruption time)** | **Y (if UE trains AI models with different capabilities based on the same data collection related configuration(s) and/or indications(s))** | **Y (if UE trains AI models with different capabilities based on the same dataset)** | **Y (if NW transfers to UE the AI models with different capabilities)** | **Y (if the reference models with different capabilities are pre-defined)** |

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | - The row of "can ensure consistency between training and inference (i.e., network additional alignment)" would not be required as the purpose of model identification itself is to allow this.  - The row of "can facilitate two-sided model pairing" is the result of Proposal 2.1.4. No duplicated discussion is required.  - The row of "is along with Model transfer" is clear from the definition of MI-Options.  - On the row of different AI model with different capabilities, we proposed following modifications.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Purpose/usage | MI-Option 1 | MI-Option 2 | MI-Option 3 | MI-Option 4 | | ~~Can ensure consistency between training and inference (i.e., network additional alignment)~~ | ~~Y~~ | ~~Y~~ | ~~Y~~ | ~~N~~ | | Can support scenario/site/cell specific model | Y | Y | Y | N | | ~~Can facilitate two-sided model pairing~~ | ~~N~~ | ~~Y~~ | ~~Y~~ | ~~Y~~ | | ~~Is along with Model transfer~~ | ~~N~~ | ~~N~~ | ~~Y~~ | ~~N~~ | | Can support different AI model with different capabilities (e.g., different performance, different interruption time) | Y (if UE trains AI models with different capabilities based on the same data collection related configuration(s) and/or indications(s)) | Y (if UE trains AI models with different capabilities based on the same dataset) | Y (if NW transfers to UE the AI models with different capabilities or UE side compile the transferred model) | Y (if the reference models with different capabilities are pre-defined or UE side compile/adjust the model with satisfying the requirement) | |
| Samsung | Some companies interpreted the agreement in RAN1#116 as the 5 options apply to Model identification Type A too. In our view, the 5 Options are for Type B.  **Regarding the purpose/usage of different options for Type B model identification, we have the following observations:**  Moreover, we think the issue here is whether there is one-to-one or one-to-many relationship between the UE’s identified model and the network-side indication. |
| ZTE | First of all, companies need to discuss the criteria to compare these different model identification options.  Regarding “Can ensure consistency between training and inference” and “Can support scenario/site/cell specific model”, it seems these two are related and can be merged. Scenario/site/cell specific model basically means there is Scenario/site/cell specific additional condition.   * Regarding whether MI-Option 4 can handle additional condition or cell/site specific model, we think it depends on the details of MI-Option 4. For example, if reference encoder is standardized, the network can apply the cell/site specific dataset collected in its cell to train the decoder. In this sense, the additional condition or cell/site specific model can be addressed. Thus, we propose to change it to “Depending on the reference model design, yes if reference encoder is standardized”.   Regarding “Is along with Model transfer”, although we can live with it, but we think it is not a necessary criterion to compare these model identification options.  Regarding “Can support different AI model with different capabilities”, we are open to discuss this.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Purpose/usage | MI-Option 1 | MI-Option 2 | MI-Option 3 | MI-Option 4 | | Can ensure consistency between training and inference (i.e., network additional alignment)  Can support scenario/site/cell specific model | Y | Y | Y | **Depending on the reference model design, yes if reference encoder is standardized** | | ~~Can support scenario/site/cell specific model~~ | ~~Y~~ | ~~Y~~ | ~~Y~~ | ~~N~~ | | Can facilitate two-sided model pairing | N | Y | Y | Y | | ~~Is along with Model transfer~~ | ~~N~~ | ~~N~~ | ~~Y~~ | ~~N~~ | | Can support different AI model with different capabilities (e.g., different performance, different interruption time) | Y (if UE trains AI models with different capabilities based on the same data collection related configuration(s) and/or indications(s)) | Y (if UE trains AI models with different capabilities based on the same dataset) | Y (if NW transfers to UE the AI models with different capabilities) | Y (if the reference models with different capabilities are pre-defined) | |
| Lenovo | We suggest the following modification.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Purpose/usage** | **MI-Option 1** | **MI-Option 2** | **MI-Option 3** | **MI-Option 4** | | **Can ensure consistency between training and inference (i.e., network additional alignment)** | **Y** | **Y** | **Y** | **N** | | **Can support scenario/site/cell specific model** | **Y** | **Y** | **Y** | **N** | | **Can facilitate two-sided model pairing** | **~~N~~ Y** | **Y** | **Y** | **Y** | | **Is along with Model transfer** | **N** | **N** | **Y** | **N** | | **Can support different AI model with different capabilities (e.g., different performance, different interruption time)** | **Y ~~(if UE trains AI models with different capabilities based on the same data collection related configuration(s) and/or indications(s))~~** | **Y (if UE trains AI models with different capabilities based on the same dataset)** | **Y (if NW transfers to UE the AI models with different capabilities)** | **Y (if the reference models with different capabilities are pre-defined)** | |
| Huawei, HiSilicon | For MI-Option 1, as clarified in Proposal 2.1.2, whether it belongs to model identification is not clear. We may use “Mechanism#1” instead.  For “support scenario/site/cell specific model”, as “Address the issue of mismatch between NW side and UE side on the categorization rule to map additional conditions into separate IDs” mentioned in Proposal 2.1.2 is not clarified, this can be marked as FFS.  For MI-Option 4, we may start discussing it after it is clarified whether it belongs to model identification.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Purpose/usage | ~~MI-Option 1~~  Mechanism#1 | MI-Option 2 | MI-Option 3 | ~~MI-Option 4~~ | | Can ensure consistency between training and inference (i.e., network additional alignment) | Y | Y | Y | ~~N~~ | | Can support scenario/site/cell specific model | FFS ~~Y~~. | Y | Y | ~~N~~ | | Can facilitate two-sided model pairing | FFS ~~N~~ | Y | Y | ~~Y~~ | | Is along with Model transfer | N | N | Y | ~~N~~ | | Can support different AI model with different capabilities (e.g., different performance, different interruption time) | Y (if UE trains AI models with different capabilities based on the same data collection related configuration(s) and/or indications(s)) | Y (if UE trains AI models with different capabilities based on the same dataset) | Y (if NW transfers to UE the AI models with different capabilities) | ~~Y (if the reference models with different capabilities are pre-defined)~~ | |
|  |  |
| Xiaomi | * For the purpose of “Can support scenario/site/cell specific model”, it seems one specific case of “Can ensure consistency between training and inference (i.e., network additional alignment)”, only keep the first Purpose seems OK. * For Option 4, we think more discussion is needed * In general, we think this proposal can be deprioritized until the procedure for each model identification type become clear. |
| Fujitsu | Support the revisions from Panasonic. |
| Nokia | Identifying main discussion points are fine. Yes, No can be discussed later. |
| NEC | We support the Panasonic recommendations |

#### **Proposal 2.1.7**

In order to reduce the scope of further study, Proposal 2.1.7 is suggested a starting point.

**Proposal 2.1.7**

**Regarding model identification Type B,**

* **Further study MI-Option 1 (including the necessity) for the use cases of AI-based positioning, AI-based beam management** 
  + **MI-Option 2/3/4 are not considered for the above use cases in Rel-19**
* **Further study MI-Option 1 (including the necessity) for the use case of CSI prediction** 
  + **MI-Option 2/3/4 are not considered for the above use case in Rel-19**
* **Further study MI-Option 2/3/4 (including the necessity) for the use case of CSI compression** 
  + **MI-Option 1 are not considered for the above use case in Rel-19**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | For site/cell specific model, to train site/cell specific model by UE side for any possible site/cell would be rather impossible. For such case, I think MI-Option 3 would be more suitable option. Either of the following should be added.  - Site/cell specific model can be considered for MI-Option 3 in Rel-19.  - Site/cell specific model is not considered in Rel-19. |
| Samsung | Ok. |
| NTT DOCOMO | We are fine with this direction at least in Rel-19. |
| New H3C | OK in general |
| ZTE | We are ok with the last bullet. However, as we discussed in the above proposals, MI-Option1 is used for addressing the additional condition issue instead of model identification. The additional condition is under discussion in each use case, and actually there is already some parallel discussion in AI Beam and AI Pos agenda, e.g., association ID in AI beam. We should avoid parallel discussion. Furthermore, it is better to first agree on the concrete additional conditions first before digging into the detailed mechanism to address the additional condition issue.  Thus, we have the following proposal.  **Regarding model identification Type B,**   * **~~Further study MI-Option 1 (including the necessity) for the use cases of AI-based positioning, AI-based beam management~~**    + **~~MI-Option 2/3/4 are not considered for the above use cases in Rel-19~~** * **~~Further study MI-Option 1 (including the necessity) for the use case of CSI prediction~~**    + **~~MI-Option 2/3/4 are not considered for the above use case in Rel-19~~** * **Further study MI-Option 2/3/4 (including the necessity) for the use case of CSI compression**    + **MI-Option 1 are not considered for the above use case in Rel-19** |
| Lenovo | We also believe MI-Option 1 can be used for two-sided cases as well, so we suggest:   * **Further study MI-Option 1/2/3/4 (including the necessity) for the use case of CSI compression** |
| Huawei, HiSilicon | We already have 2.1.4 to discuss the applicable cases for MI-Options. Why do we have a redundant proposal for 2.1.7?   |  | | --- | | **Proposal 2.1.4**  **Regarding model identification type B,**   * **MI-Option 1 can be used for UE-sided model** * **MI-Option 2 can be used for one-sided model and two-sided model** * **MI-Option 3 can be used for UE-sided model and two-sided model** |   Our view is the same, that MI-Option 1 for one-sided model can be FFS after whether it belongs to model identification is clarified. For MI-Option 2/3, they at least apply to two-sided model.  Regarding model identification Type B,   * ~~Further study MI-Option 1 (including the necessity) for the use cases of AI-based positioning, AI-based beam management~~    + ~~MI-Option 2/3/4 are not considered for the above use cases in Rel-19~~ * ~~Further study MI-Option 1 (including the necessity) for the use case of CSI prediction~~    + ~~MI-Option 2/3/4 are not considered for the above use case in Rel-19~~ * Further study MI-Option 2/3~~/4~~ (including the necessity) for the use case of CSI compression   FFS MI-Option 1 ~~are not considered~~ for the above use case in Rel-19 |
| Xiaomi | For MI-Option 3, it is related to the model transfer. Whether support model transfer for one sided model is one separate discussion. It should be discussed within the model transfer/delivery scope and can’t be precluded in this proposal. |
| QC | OK |
| Nokia | Ok with the direction. |
| NEC | We also share the similar question on the relationship between Proposal 2.1.4 and this proposal. |

# Training data collection for UE-sided model

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Huawei[1] | *Proposal 5: Data categorization indication for UE side data collection of one-sided model, if needed, could be studied with local ID rather than globally unique ID.*  *Proposal 13: 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.* |
| Ericsson[3] | *Proposal 12 Conclude that the Rel-18 LS response to RAN2 is sufficient for addressing the study objective on data content, at least as a starting point for Rel-19.*  *• RAN2 can send an LS to RAN1 if there is a need to discuss any additional content, or any further details of the content.* |
| Spreadtrum[4] | *Proposal 1: For data collection for UE-side model training, support 1a or we could wait the progress of RAN2.* |
| Intel[5] | *Proposal 7:*  *• On CN/OAM/OTT collection of UE-sided model training data, RAN1 to consider further on the following aspects:*  *o Necessity of 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.*  *o Details of contents of the collected data, considering the details listed in R1-2310681 as a starting point.* |
| Vivo[6] | *Proposal 11: For CN/OAM/OTT collections of UE-sided model training data, the data content can refer to the agreed LS table in [3][4].*  *Observation 12: Information of relationship of Set A and Set B on beam width/beam pointing angle/beam pattern would be needed for AI based beam management.*  *Observation 13: PRS configuration and Tx beam related information would be needed for AI based positioning.*  *Proposal 12: Additional information for UE side data collection include NW configuration information and Tx beam related information.*  *Proposal 13: Implicit method (dataset categorization) and explicit method (explicitly indicated assistance information) could be used for provision of Tx beam related information.* |
| ZTE[7] | *Proposal 7: 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.* |
| Google[8] | *Proposal 4: Support the NW and UE to maintain the same understanding on when the UE can perform data collection.*  *Proposal 5: Support the NW to provide side information to facilitate the UE data collection.* |
| OPPO[9] | *Proposal 7: 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.* |
| CATT[10] | *Observation 11: 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 9: 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.* |
| Samsung[11] | *Observation#1: 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#2: 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#6: 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.* |
| CMCC[12] | *Proposal 7: Regarding the UE side data collection mechanism, RAN2 could take the Reply LS on Data Collection Requirements and Assumptions (R1-2310681) as the baseline.* |
| LGE[13] | *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.* |
| Xiaomi[14] | *Proposal 8: The data content and related information included in RAN1 LS (R1-2310681) to RAN2 can be set as baseline* |
| Fujitsu[17] | *Proposal-7: From the RAN1 perspective, the focus of the study on the collection of UE-sided model training data is on identifying the corresponding contents of UE data collection. The continued study on 7.2.1.3.2 is left to RAN2.*  *Observation-5: Assistance information can be used to convey the additional condition in data collection.*  *Proposal-8: Assistance information for data categorization needs to be further studied.*  *Proposal-9: Both NW-side additional conditions and UE-side additional conditions can be taken as assistance information in data categorization and can be for further study.*  *Proposal-10: For the details of assistance information, we suggest that*  *• Which aspects/details can be considered as additional condition/assistance information is left to per-use-case study*  *• Which aspects belong to proprietary information and how to avoid the disclosure of proprietary information can be studied together*  *Proposal-11: The following aspects are suggested to be studied to facilitate the differentiation/categorization of training data.*  *• Assistance information required to check model generalization capability*  *• Assistance information required to differentiate cell/scenario/area for local models*  *• MI-Option 1 related data indications and configurations*  *• Other aspects if any*  *Proposal-12: The quantization of data samples in data collection needs to be studied.*  *Proposal-13: Regarding the ground truth label in data collection, the following aspects are suggested to be studied:*  *• Availability of the ground truth label*  *• Quality of the ground truth label*  *• Quantization of the ground truth label* |
| IDC[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 3: 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 4: 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 5: Quality indicators for the CSI prediction model training dataset could include at least the RSRP and TDCP.*  *Proposal 6: A ground truth label quality indicator is associated with a UE or PRU location*  *Proposal 7: Support both hard (1 or 0) and soft indicator (0, 0.1, 0.2, …, 1.0) for a ground truth label quality indicator*  *Proposal 8: 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 9: 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 10: The LMF is the only entity that can generate a ground truth label quality indicator associated with location information*  *Proposal 11: 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 12: 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 13: 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.* |
| NVIDIA[21] | *Proposal 3: Conclude that there is a need for collection of UE-sided model training data.* |
| Apple[22] | *Proposal 2: From RAN1 perspective, option 1-1a is sufficient and no additional requirement is identified to enhance data collection for UE side model training.* |
| Nokia[24] | *Observation 13: Enabling data collection (identification of triggering and terminating NW entity, procedural and signalling details) for UE-sided model training go beyond RAN1 realm.*  *Proposal 7: RAN1 to consider following priority when studying the listed objectives of this sub-agenda,*  *• CN/OAM/OTT collection of UE-sided model training data collection*  *o Study contents, type and format of training data based on use case requirements*  *o Study necessity of assistance information for categorizing the training data*  *Proposal 8: For CN/OAM/OTT collection of UE-sided model training data, RAN1 to clarify enabling of which data measurement or data configuration require involvement of CN/OTT/OAM.*  *Proposal 9: For CN/OAM/OTT collection of UE-sided model training data, RAN1 to agree that enabling data collection (identification of triggering and terminating NW entity, signalling details) for UE-sided model training goes beyond RAN1 (and RAN WGs).* |
| ETRI[25] | *Proposal 4: Datasets should be categorized based on NW configurations and configured functionalities during the data collection process.*  *Proposal 5: The NW can request UEs to transfer collected data immediately for the purpose of categorizing the dataset.*  *Proposal 6: The UE needs a mechanism to categorize data samples according to changes in its settings.* |
| Qualcomm[27] | *Proposal 11: 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 OTT server (in a 3GPP transparent or 3GPP non-transparent method)*  *- Transfer of the collected data to the OTT server (via CN or OAM).*  *Observation 6: The actual input/output and side/auxiliary information for a UE-side model are implementations-specific choices and cannot be pre-determined/standardized.*  *Observation 7: 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, or third parties.*  *Proposal 12: 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 8: 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 13: Considering the implementation-specific nature of the model input/output and auxiliary/side information and considering the runtime constraints (as mentioned in observation 8), the UE-side model can only be trained by the UE vendor, at least in the Rel-19 and foreseeable near future.* |
| DCM[28] | *Proposal 5: When considering data collection toward OTT server, the ownership of data should be clarified first.* |
|  |  |
|  |  |

#### **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, more than 10 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.

In the last meeting, we had a proposal to study the assistance information from NW to UE to facilitate the categorization of training data. However, most companies thought there were nothing beyond what we were discussing or what we had discussed before.

## 1st round discussion

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:

|  |  |
| --- | --- |
| Huawei[1] | *Observation 5: For model transfer/delivery where the model is trained at NW side, the feasibility for Case z2 is unclear since it may incur the burden of offline cross-vendor collaboration.*  *Observation 6: 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 7: For model transfer/delivery where the model is trained at UE side or neutral site, the necessity of introducing Case z1 and Case z3 as opposed to the implementation manner of Case y is not clear:*   * *Case z1 and Case z3 incur the burden of offline cross-vendor collaboration, compared to Case y.* * *Case z1 and Case z3 may come with 3GPP NW side burden on model maintenance/storage compared to Case y.* * *Case z1 and Case z3 do not bring benefits compared to Case y.*   *Proposal 12: For model transfer/delivery where the model is trained at UE side or neutral site, assume Case y as the baseline.* |
| Ericsson[3] | *Proposal 13 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 14 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.* |
| Spreadtrum[4] | *Proposal 2: Suggest to defer the discussion on model transfer/delivery until good progress on AI9.1.3.2 multi-vendor issue achieved.* |
| Intel[5] | *Observation 3:*  *• 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.*  *Proposal 5:*  *• In Rel-19, consider support of collaboration level y from the perspective of 3GPP specifications and continue discussions on additional support of other collaboration levels under level z.*  *Observation 4:*  *• 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 6:*  *• For model transfer in a proprietary format, support a NW to update the parameters of a model at the UE, including indirect means involving a UE-side server if needed.* |
| Vivo[6] | *Proposal 6: 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 5: The burden of model storage would be relieved if the model structure is specified in 3GPP.*  *Observation 6: Proprietary design disclosure may not be a concern if the model structure is widely known and does not involve any device-specific design decisions.*  *Observation 7: From initial results for field test, cell/site specific model can provide up to 17.6% SGCS gain.*  *Observation 8: Field test shows model developed for one cell does not perform well for the other cell.*  *Observation 9: Field test shows that simple and small models work well for different cases, at least for cell/site specific model.*  *Observation 10: SF compression with cell/site specific model provides up to 11.4% gain compared to legacy codebook in Uma scenario with 0% indoor UE distribution.*  *Observation 11: For UMa dense urban scenario, TSF compression with cell/site model achieves additional gain compared to the general model, with the maximum gain up to 12.5% for case2 and 6.6% for case3.*  *Proposal 7: Defining reference model (structures) is also beneficial from RAN4 testing perspective.*  *Proposal 8: Support model transfer with known model structure at UE (Case z4).*  *Proposal 9: 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 10: Further study necessity of post deployment testing. Take CSI compression as example, several monitoring methods can be considered as options for post deployment testing:*  * Method 1: The UE obtains a test dataset containing only the channel (encoder input), obtains the PMI by encoder inference, and then reports the PMI to the NW. NW decide the results.*  * Method 2: UE obtains the test dataset containing both the channel and PMI. UE decide the results.*  * Method 3: UE reports the channel and PMI to NW. NW decide the results.* |
| ZTE[7] | *Proposal 5: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer study for two-sided model rather than UE-side model.*  *Proposal 6: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer z4 for two-sided model.* |
| OPPO[9] | *Proposal 8: To consider the necessity of the standardized model transfer/delivery solutions, a comparison between 3GPP-standardized solution and non-3GPP solution is needed, for resolving the burden of offline cross-vendor collaboration, burden on model maintenance/storage, proprietary design disclosure concern.* |
| CATT[10] | *Observation 9: For one-sided model at UE side, if the model is trained by UE side (UE vendor) itself, there is no need to pursue standardized solution for model transfer/delivery.*  *Observation 10: For UE-side models that trained by NW-side or two-sided models, it may be beneficial to consider standardized solution for model transfer/delivery to alleviate inter-vendor coordination effort.*  *Proposal 7: No need to pursue model transfer case z1, z2 or z3 in Rel-19.*  *Proposal 8: Further study model transfer z4 in Rel-19.*  *- As a starting point, a small set of simple model structures can be considered as reference model structures;*  *- Further study the value of z4 in terms of reduced effort from inter-vendor offline coordination, compared to case y with NW-side training;*  *- Further study the feasibility of parameter update with the same model structure in UE device.* |
| Samsung[11] | *Proposal#7: Deprioritize study on 3GPP non-transparent model transfer cases that require 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#8: Study the feasibility and potential benefits of model (parameter) transfer for specified model structure from gNB to UE, i.e., Case z4.* |
| CMCC[12] | *Proposal 5: 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 6: For each case of model transfer/delivery to UE, the following aspects can be discussed further:*  * The detailed components of model, including open format and proprietary format*  * The NW/UE requirements (e.g., model compiling capability) before model transfer/delivery*  * The necessary components of model during model transfer/delivery*  * The transfer /delivery container and corresponding requirements (e.g., model size) during model transfer/delivery*  * The transfer/delivery latency if model need update/retrain/finetune*  * The deployment delay, corresponding procedures after model transfer/delivery* |
| LGE[13] | *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.* |
| Xiaomi[14] | *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: In case y, when the AI models are developed by network, potential specification effort on the assistance signalling/procedure for the model transfer/delivery is necessary*  *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 for case y, case z2 and case z4* |
| NEC[16] | *Observation 4: Supporting model transfer is essential when considering cell/scenario-specific AI/ML deployment which is expected to happen when AI/ML deployment accelerates.*  *Proposal 12: Model transfer should be supported from Rel-19 to ensure future-proofness of AI/ML operation.*  *Observation 5: Model transfer methodology z2 requires complex UE-network collaboration which increases network complexity significantly.*  *Observation 6: Model transfer in open format reduces complexity of inter-vendor UE-network collaboration which can be helpful for use cases like CSI compression.*  *Proposal 13: Prioritize z1, z3 and z4 for further discussion for model transfer methodologies.*  *Proposal 14: Support group common transmission for model transfer.* |
| Fujitsu[17] | *Observation-6: From the RAN1 perspective, a further study on model transfer/deliver depends on the progress and conclusions on cell/site-specific model and training data collection of UE-side model.*  *Proposal-14: From the RAN1 perspective, it can be considered to defer the study on model transfer/delivery until the following issues are clarified:*  *• The benefits of using cell/site-specific model*  *• Training data collection including data categorization for UE-side model* |
| MediaTek[18] | *Observation 2-1: Instead of arguing the need of model ID after model identification, the need and the benefit of model identification may be more critical for exploration*  *Observation 2-2: For ensuring consistency, the option to provide information related to additional conditions is equivalent to check the information at training stage and at inference stage respectively. If they are aligned, there is consistency*  *Observation 2-3: When the model is built at the training stage, it still has the risk whether the model could be used for inference, since there is uncertainty on whether there is change on the NW side configurations/conditions after model development*  *Observation 2-4: It would be expected that the model identification method could provide the way to eliminate the uncertainty at NW and UE part during model development. Namely, once a model is identified, it guarantees the consistency of the AI/ML model for inference in the future. Otherwise, the provision of assistance data related to configurations and conditions at different stages for checking the alignment is enough*  *Observation 2-5: Model ID could be a field within an IE with multiple fields that describes a model’s property after model identification*  *Proposal 2-1: Strive to develop model identification method to ensure NW side consistency between training stage and the future inference stage, not just to allocate model ID*  *The below is a brief summary for our views,*  *• For case y, UE has largest degree of freedom for AI/ML framework design*  *• For case z1, UE has benefit to reduce proprietary signalling design effort*  *• For case z2, this may not be a good use case*  *• For case z3, it could be treated as a temporary solution for moving to case z1 in the future*  *• For case z4, it can’t show the UE differentiation*  *Proposal 3-1: Further study on case z2 and z4 whether one of them could be further deprioritized* |
| IDC[20] | *Observation 11: In cases where model generalization, model finetuning or model storage/switching is not feasible, model delivery/transfer can be beneficial.*  *Proposal 14: Model transfer for UE-side models with functionality-based LCM is not supported and 3GPP specification transparent model delivery is only considered.* |
| NVIDIA[21] | *Proposal 4: Conclude that there is a need to consider standardised solutions for transferring/delivering AI/ML model(s).* |
| Apple[22] | *Proposal 3: The necessity of standardized model transfer solution for case z1-z5 can be further discussed after CSI compression use case is concluded.* |
| Nokia[24] | *Observation 10: Discussions and solutions on how UE-side ML models are made available in their respective training location (particularly, if in 3GPP NW) are not in the scope of 3GPP RAN work.*  *Observation 11: RAN1 can potentially address jointly, based on use case requirements, the need for collaboration levels z1 and z3.*  *Proposal 4: Clarify based on use case requirements, the need for the collaboration levels z4 within the scope of Release 19 specifications.*  *Proposal 5: As there are no use case requirements, RAN1 to not specify model transfer/delivery solutions z1-z4 for UE-sided beam management and positioning accuracy enhancement use cases.*  *Observation 12: For 2-sided CSI compression, particularly training type I (joint model training and model transfer/delivery to the UE), model transfer can be realized as user plane data transfer, controlled by the gNB/RAN.*  *Proposal 6: RAN1 to study the potential model transfer/delivery solutions for 2-sided CSI feedback enhancement (particularly for training type I) based on their performance requirements.* |
| AT&T[26] | *Proposal 8: Model transfer/delivery is supported for both UE-sided models and UE-part of two-sided models in Rel-18.*  *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 9: 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 10: 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, z2) 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 z3, 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.* |
| Qualcomm[27] | *Proposal 15: Add the following case to the existing model transfer/delivery cases:*  *• Case z0 (or y2) [not obvious which case it belongs to, hence the naming]: model is trained at UE-side (i.e., at a UE-side OTT server) and the model transfer happens from the OTT server to UE in a non-transparent manner.*   |  |  |  |  | | --- | --- | --- | --- | | **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 | | **z0 (or y2)** | **model transfer (if needed) over-the-top.** | **UE-side (including UE-side OTT server hosted at 3GPP Network)** | **UE-side** | | **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 | |
| DCM[28] | **Table 3.**  Characteristic of model delivery/transfer for one sided model, where the model is trained at UE side.   |  |  |  |  | | --- | --- | --- | --- | | Model transfer/deliver case | Case y (UE training) | Case z1 | Case z3 | | Friendly to UE’s implementation | Friendlier as the model can be offline compiled and tested | Friendlier as the model can be offline compiled and tested | Less friendly if UE need to compile/test the model  Friendly same as case y and case z1 otherwise | | Burden of offline collaboration between NW and UE/chipset vendors | Heavy burden | Heavy burden | Heavy burden | | Burden on the training data collection at NW ~~at 3GPP network~~ | No | No | No | | Burden on the model storage and model management for model transfer/delivery at NW ~~at 3GPP network~~ | No | Yes | Yes | | Risk of proprietary design disclosure | No | Higher risk compared to Case y | Higher risk compared to Case y, z1 | | Potential spec impact for model transfer/delivery | No spec impact on 3GPP air interface | More spec impact on 3GPP air interface compared to Case y | More spec impact on 3GPP air interface compared to Case y, z1 | | ~~Support of site/cell-specific model~~ |  |  |  |   **Table 4.**  Characteristic of model delivery/transfer for one sided model, where the model is trained at NW side.   |  |  |  |  | | --- | --- | --- | --- | | Model transfer/deliver case | Case y (NW training) | Case z2 | Case z4 | | Friendly to UE’s implementation | Friendlier as the model can be offline compiled and tested | Friendlier as the model can be offline compiled and tested | Less friendly if UE need to compile/test the model compared to Case y;  Friendly same as case y and case z2 otherwise | | Burden of offline collaboration between NW and UE/chipset vendors | Heavy burden | Heavy burden | Less burden\* compared to Case y, z2 | | Burden on the training data collection at NW ~~at 3GPP network~~ | Yes | Yes | Yes | | Burden on the model storage and model management for model transfer/delivery at NW ~~at 3GPP network~~ | Yes | Yes | Yes | | Risk of proprietary design disclosure | Disclosed to UE/chipset vendor | Disclosed to UE/chipset vendor | Disclosed to UE/chipset vendor | | Potential spec impact for model transfer/delivery | No spec impact on 3GPP air interface | Need of spec impact on 3GPP air interface | More spec impact on 3GPP air interface compared to Case z2 | | ~~Support of site/cell-specific model~~ |  |  |  |   **Table 5.**  Characteristic of model delivery/transfer for two sided model, where the model is trained at UE side.   |  |  |  |  | | --- | --- | --- | --- | | Model transfer/deliver case | Case y (UE training) | Case z1 | Case z3 | | Friendly to UE’s implementation | Friendlier as the model can be offline compiled and tested | Friendlier as the model can be offline compiled and tested | Less friendly if UE need to compile/test the model  Friendly same as case y and case z1 otherwise | | Burden of offline collaboration between NW and UE/chipset vendors | Heavy burden | Heavy burden | Heavy burden | | Burden on the training data collection at NW ~~at 3GPP network~~ | No | No | No | | Burden on the model storage and model management for model transfer/delivery at NW ~~at 3GPP network~~ | No | Yes | Yes | | Risk of proprietary design disclosure | No for UE-part | Higher risk for UE part compared to Case y | Higher risk for UE part compared to Case y, z1 | | Potential spec impact for model transfer/delivery | No spec impact on 3GPP air interface | More spec impact on 3GPP air interface compared to Case y | More spec impact on 3GPP air interface compared to Case y, z1 | | Additional method for Pairing of UE-part and NW-part of two-sided model | May Needed | May Needed | May Needed | | ~~Support of site/cell-specific model~~ |  |  |  |   **Table 6.**  Characteristic of model delivery/transfer for two sided model, where the model is trained at NW side.   |  |  |  |  | | --- | --- | --- | --- | | Model transfer/deliver case | Case (NW training) | Case z2 | Case z4 | | Friendly to UE’s implementation | Friendlier as the model can be offline compiled and tested | Friendlier as the model can be offline compiled and tested | Less friendly if UE need to compile/test the model compared to Case y;  Friendly same as case y and case z2 otherwise | | Burden of offline collaboration between NW and UE/chipset vendors | Heavy burden | Heavy burden | Less burden compared to Case y, z2 | | Burden on the training data collection at NW ~~at 3GPP network~~ | Yes | Yes | Yes | | Burden on the model storage and model management for model transfer/delivery at NW ~~at 3GPP network~~ | Yes | Yes | Yes | | Risk of proprietary design disclosure | UE-part disclosed to UE/chipset vendor | UE-part disclosed UE/chipset vendor | UE-part disclosed UE/chipset vendor | | Potential spec impact for model transfer/delivery | No spec impact on 3GPP air interface | Need of spec impact on 3GPP air interface | More spec impact on 3GPP air interface compared to Case z2 | | Additional method for Pairing of UE-part and NW-part of two-sided model | Not needed | Not needed | Not needed | | ~~Support of site/cell-specific model~~ |  |  |  |   *Proposal 6: In the model delivery/transfer discussion, it should be avoided to ignore the NW side burden that is transparent in 3GPP specification.*  *Proposal 7: Remove the row “Support of site/cell specific model” from the table.* |
|  |  |
|  |  |

#### **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. | | | |

## 1st round discussion

#### **Proposal 4.1.1**

Based on the tdocs, most companies think Case z1 and z3 are less attractive compared to Case y due to various reasons. Thus, Proposal 4.1.1 is suggested for further discussion.

**Proposal 4.1.1**

**From RAN1 perspective, the model transfer/delivery Case z1 and z3 are 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**

Companies can provide comments/inputs in the following table:

|  |  |  |
| --- | --- | --- |
| Company | Comment | |
| Panasonic | I have the concern that the transfer cost of case y would be charged to the terminal customer for case y but it is beyond RAN1 expertise. As far as "from RAN1 perspective" is kept, I'm ok. | |
| Samsung | Thank you FL. It is Ok for us. | |
| NTT DOCOMO | Support the proposal. | |
| New H3C | OK in general | |
| ZTE | Support the proposal. | |
| Huawei, HiSilicon | Support. | |
| Fujitsu | Support the proposal | |
| vivo | We prefer to at least keep z3, which is still under discussion in two sided CSI compression use case. | |
| QC | First, we suggest differentiating UE-side models from NW-side models in the discussions related to model transfer/delivery cases and the corresponding prioritization discussions. For UE-side models, we should first focus on deprioritizing cases z2 and z4, due to major feasibility issues associated with training UE-sided models at the NW side, including, but not limited to the issues with regards to testability, efficient implementations related to target UEs, etc.  Additionally, Case z1 is being actively pursued and discussed in RAN2, so we should not consider deprioritizing it in RAN1 at this stage. | |
| mtk | 1, no need to deprioritize z1 at this moment. For z1, the storage is at 3GPP network. We consider the use case that the UE vendor could lease the OTT server from the operator within 3GPP NW, so that it is still like UE vendor to have model storage within 3GPP NW. Maybe our use case is similar to the new one proposed by QC |
| Nokia | Ok | |
| NEC | We think z3 can address issues related to inter-vendor UE-network collaboration. For instance, supporting standardized reference model structure or model parameters exchange between UE and network is currently considered as the potential solutions to address the complexity of UE-network collaboration for CSI compression use case. | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |

#### **Proposal 4.1.2**

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, Proposal 4.1.2 is suggested to restrict the scope of Case z4.

**Proposal 4.1.2**

**For model transfer/delivery Case z4, Rel-19 study focuses on the option with standardized known model structure(s).**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | Support |
| Samsung | Ok |
| NTT DOCOMO | Fine with the proposal. |
| New H3C | OK in general |
| ZTE | Support |
| Huawei, HiSilicon | OK |
| Fujitsu | Support |
| vivo | Support |
| QC | Not support. It is not clear why we should only focus on "standardized" known model structure? What about unstandardized known model structure? This is an active discussion in 9.1.3.2 (Inter-vendor collaboration discussion Option 3 vs. Option 5), so we should not deprioritize non-standardized structures at this stage. |
| Nokia | OK |
| NEC | Support this direction. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 4.1.3**

Regarding Case z4, one of the key potential advantages is to reduce the offline collaboration between different vendors. However, some companies suggest some requirement, which still requires offline cross-vendor collaboration, for Case z4, e.g., testing of the parameters. In contrast, some companies have the opposite view. Thus, Proposal 4.1.3 is suggested in order to have a common understanding on the potential requirement(s) for Case z4.

**Proposal 4.1.3**

**From RAN1 perspective, the following requirements for model transfer/delivery Case z4 are needed:**

* **The structure of AI model should be tested before the model transfer in order to ensure the performance and UE implementation**
* **[The parameters of AI model should be tested before the model transfer in order to ensure the performance and UE implementation]**
* **…**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | Even non-AI/ML functionality case, inter-operability test is realized by the offline collaboration between different vendors.  Test is essential but when and how tests are carried out should be discussed separately. Both of followings are reasonable. The need of the first bullet corresponds to the need of offline test and it can be discussed later.  - Before model transfer of commercial deployment, offline test is carried out between different vendors.  - Before model transfer of commercial deployment, model transfer with small scale test operation is carried out for the test/validation of the model. |
| Samsung | This can be studied first. |
| ZTE | The first bullet needs to be clarified first. There are at least the following two different understandings.  Understanding#1: The structure of AI model should be tested before entering the market.  Understanding#2: The structure of AI model should be tested before base station transfers model to the UE **every time**.  If the intention of this proposal is understanding#1, we can agree with the intention. However, we don’t see the need for understanding#2. |
| Huawei, HiSilicon | For whether the exchanged parameters has to be tested at OTT server or run at UE with plug-and-play manner, both directions can be studied and clarified first. |
| Fujitsu | If the model structure in Z4 is “**with standardized known model structure(s)**”, the model structure is known at UE-side. Then the test is up to UE implementation.  Regarding model parameter transfer, the test is needed from UE implementation perspective. But it would be difficult to do such test at NW-side. |
| vivo | We don’t think the second bullet is needed since model structure test itself should already serve the purpose of testing whether parameter update functionality at UE side fulfill the requirements.  If we want to study this, RAN4 may also need to be involved. |
| QC | The model (including its structure **and** parameters) should have been fully tested for the target UEs, and its support should have been indicated by the UE capability, for all model transfer/delivery cases, including z4. |
| Nokia | Not clear about the need of this proposal. |
| NEC | Support in principle but we are not clear about what is the spec impact, is it intended to introduce a dedicated validation procedure? |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 4.1.4**

Based on the tdocs, most companies are not supportive of Case z2. In contrast, it seems only one company supports Case z2 and another company says Case Z2 is feasible. Thus, Proposal 4.1.4 is suggested to check whether we can deprioritize z2 or not now.

**Proposal 4.1.4**

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

* **Risk of proprietary design disclosure**
* **Large burden of offline cross-vendor collaboration**

Companies can provide comments/inputs in the following table:

|  |  |  |
| --- | --- | --- |
| Company | Comment | |
| Panasonic | Support | |
| Samsung | Ok | |
| NTT DOCOMO | Support the proposal. | |
| New H3C | OK in general | |
| ZTE | OK | |
| Huawei, HiSilicon | Support | |
| Fujitsu | Support | |
| QC | OK for UE-sided models. For two-sided models we should wait for the discussion and conclusion form 9.1.3.2, as different inter-vendor collaboration options are being actively discussed. | |
| mtk | ok |
| Nokia | OK | |
| NEC | OK | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |

#### **Proposal 4.1.5**

Qualcomm proposes a new case that was not been discussed in R18 SI. It is better to check companies’ view on whether further study is needed or not.

**Checkpoint 4.1.5**

|  |  |  |  |
| --- | --- | --- | --- |
| **Case** | **Model delivery/transfer** | **Model storage location** | **Training location** |
| **z0 (or y2)** | **model transfer (if needed) over-the-top.** | **UE-side (including UE-side OTT server hosted at 3GPP Network)** | **UE-side** |

Companies can provide comments/inputs in the following table:

|  |  |  |
| --- | --- | --- |
| Company | Comment | |
| Mod | * Not a proposal. Just to check companies’ view on whether further study on z0/y2 is needed or not. * @QC: Would you like to clarify whether there is any spec impact on the air interface or not? It is also better to clarify the procedure and/or impact within 3GPP so that other companies can have a better understanding | |
| Panasonic | It is not yet clear the difference from case y. UE-side is not outside 3GPP network? | |
| Samsung | Same view as moderator. Clarification is needed. | |
| New H3C | OK in general | |
| ZTE | Same view as moderator. Clarification is needed. | |
| Huawei, HiSilicon | Agree to clarify first. BTW, if any 3GPP is involved in the procedure, this behaviour should be categorized as Case z1 (since 3GPP network is to store the model)? | |
| Fujitsu | Same view as moderator. Clarification is needed. | |
| QC | In response to the moderator and other companies asking for clarification: For this new option, there will be specification impact that are beyond RAN1/RAN2. There may be air interface changes (such as special handling or scheduling) but it is beyond RAN1 and it is a common discussion with other model delivery/transfer options. RAN1 is involved if the models are specified (z3,z4,z5). This option should be included as it is related to the data collection discussion in RAN2 option 1b, where the data is collected in a vendor server within the operator network. Then a model can be trained within an operator network and downloaded back to the UEs. Citing the related RAN2 agreement here for reference:  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. UE collects and directly transfers training data to the Over-The-Top (OTT) server; 1a) OTT (3GPP transparent)  1b) OTT (non-3GPP transparent) UE collects training data and transfers it to Core Network. Core Network transfers the training data to the OTT server.UE collects training data and transfers it to OAM. OAM transfers the needed data to the OTT server. In response to Panasonic: Case y is a baseline that does not require 3GPP discussions. The new proposed case will have impacts to 3GPP specification but may be not RAN1/RAN2 specifications. RAN/SA plenaries have opened the door for this discussion to happen in SA2. It is an option that can satisfy all requirements and should be included in our options. | |
| mtk | 1, It seems to be similar to how we treat case z1. The UE vendor leases the OTT server from the operator and it is within 3GPP NW. |
| Nokia | It is not clear why to expand model transfer discussions any further. RAN1 is not a place to discuss further details on OTT servers, etc.. | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |

# Others

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| ZTE[7] | *Proposal 1: RAN1 and RAN2 are advised to avoid repeating the Rel-18 discussions in Rel-19 and instead concentrate on addressing the remaining open issues.*  * RAN1 can focus on the requirement and necessity part, while RAN2 can focus on the detailed mechanism design*  * At least for model identification and model transfer/delivery, RAN2 can wait for the input from RAN1 first before proceeding with specifics* |
| CATT[10] | *Observation 12: For functionality-based LCM, whether and how to assess/monitor the performance of an inactive model at UE-side is up to UE implementation.*  *Proposal 10: For model-ID-based LCM (if supported), it is beneficial to assess/monitor the performance of a specific inactive model at UE-side.*  *Proposal 11: For functionality-based LCM, it is beneficial to assess/monitor the performance of a specific inactive functionality at UE-side.*  *Proposal 12: Further study the additional conditions of the following cases:*  *- Ensuring the consistency of UE-side additional condition for UE-sided model from NW perspective, if NW controls the model in model-ID-based LCM is supported.*  *- Ensuring the consistency of UE-side additional condition for NW-sided models from NW perspective.*  *- Ensuring the consistency of UE-side and NW-side additional condition, if two-side model use case (i.e. CSI compression) is supported.* |
| LGE[13] | *Observiation#1. Studies on model identification and delivery/transfer were led by RAN1 in Rel-18 but decided to switch to RAN2 in Rel-19 due to discussion inefficiency caused by lack of expertise.*  *Observiation#2. The second objective is corresponding to a remaining work in RAN2 in Rel-18.*  *Proposal#1. Work split between RAN1 and RAN2 should be clearly defined on this agenda.*  *Proposal#2. RAN1 should not work on further elaboration on types/categories on model identification and model delivery/transfer.* |
| NEC[16] | *Observation 2: Ensuring consistency of additional conditions using monitoring procedure results in high delay in identification of the suitable AI/ML model to run at UE, during which system performance suffers.*  *Proposal 6: 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 7: 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 8: 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.*  *Observation 3: Concurrent inference operation of two models/functionalities at a UE (where one model/functionality is inactive but being monitored and other model/functionality is activated at UE) allows testing of newly deployed AI/ML model/functionality (using inactive model operation) and at the same time continuing the radio operation using older well-established AI/ML model/functionality.*  *Proposal 9: 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 10: Support adaptive model/functionality selection, activation, deactivation, switching, and fallback based on additional conditions.*  *Proposal 11: Support event triggered AI/ML functionality/model activation/deactivation/switching.*  *Proposal 15: 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.* |
| Continental Automotive[19] | *Proposal 5: Mapping relation information can be configured for model versions in association with additional condition grouping or segmentation related to model training.*  *Proposal 6: Paired additional conditions on both NW and UE sides can be identified for alignment with the pre-configured information (e.g., index or ID).*  *Proposal 7: Selection of candidate inactive models need to be further studied in terms of improving model switching performance and minimizing any potential impact (e.g., signalling overhead).*  *Proposal 8: Study of online training is suggested.*  *Proposal 9: Study of UE ML capability related to training collaboration aspect is suggested.* |
| NVIDIA[21] | *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.*  *Proposal 1: Conclude that there is a need for model identification in the context of LCM.*  *Proposal 2: Studies the following options as starting point for model identification type B with more details related to all use cases:*  *• MI-Option 4: Model identification via standardization of reference models. (for CSI compression)*  *• MI-Option 5: Model identification via model monitoring* |
| AT&T[26] | *Proposal 1: Study the following aspects that are necessary for the common framework for 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 11: 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 12: 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 13: 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 14: 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.*  *o 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.*  *Proposal 15: For UE sided models and two-sided models, for models that are not transparent to the network, UE-autonomous mechanisms should not be considered for selection, activation, deactivation, switching, and fallback and the final decision should be made by the network:*  *• Decision by the network*  *o Network-initiated*  *o UE-initiated, requested to the network.*  *• Decision by the UE*  *o Event-triggered as configured by the network or predefined by spec, UE’s decision is reported to network.*  *Proposal 16: Confirm the necessity of assessment/monitoring of inactive models / functionalities, with the following assumptions as the starting point:*  *• One way to monitor inactive models/functionalities is by activating them and reusing mechanisms defined for monitoring of active models/functionalities.*  *o FFS: feasibility of activating multiple models/functionalities.*  *• The following aspects may be considered for further study or in WI to assess the applicability and expected performance of an inactive model/functionality:*  *o Configuring an AI/ML model(s) for monitoring without activation (e.g., monitoring-only mode without reporting predicted beams in BM Case 1 and 2)*  *o Dataset delivery / RS configuration from the network to the UE for assessment/monitoring of the applicability and expected performance of the model/functionality.*  *o The procedure and signaling for NW-side assessment/monitoring and UE-side assessment/monitoring.*  *o NW may provide performance criteria/preference for UE’s model selection.*  *o Other aspects are not precluded for further study or specification.*  *Target performance may be aligned during model identification, in addition to any RAN4 tests.* |
|  |  |

**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

# 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

# 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 | Hualei.wang@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 | [guan\_peng@nec.cn](mailto:guan_peng@nec.cn)  [pravjyot.deogun@EMEA.NEC.COM](mailto:pravjyot.deogun@EMEA.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 |
|  |  |  |

# Appendix B: Reference/tdocs

1. R1-2402027 Discussion on other aspects of the additional study for AI/ML Huawei, HiSilicon
2. R1-2402052 Discussion on other aspects of AI/ML model and data on AI/ML for NR air-interface FUTUREWEI
3. R1-2402057 Discussion on other aspects of AI/ML Ericsson
4. R1-2402097 Discussion on other aspects of AI/ML model and data Spreadtrum Communications
5. R1-2402148 Other study aspects of AI/ML for air interface Intel Corporation
6. R1-2402234 Other aspects of AI/ML model and data vivo
7. R1-2402267 Discussion on study for other aspects of AI/ML model and data ZTE
8. R1-2402280 AI/ML Model and Data Google
9. R1-2402320 Additional study on other aspects of AI/ML model and data OPPO
10. R1-2402370 Additional study on other aspects of AI/ML model and data CATT, CICTCI
11. R1-2402456 Discussion for further study on other aspects of AI/ML model and data Samsung
12. R1-2402557 Discussion on other aspects of AI/ML model and data CMCC
13. R1-2402631 Discussion on other aspects of AI/ML model and data LG Electronics
14. R1-2402653 Further study on AI/ML model and data Xiaomi
15. R1-2402695 Discussion on other aspects for AI/ML for air interface Panasonic
16. R1-2402757 Discussion on other aspects of AI/ML model and data NEC
17. R1-2402790 Discussion on other aspects of AI/ML model and data Fujitsu
18. R1-2402800 View on AI/ML model and data MediaTek Korea Inc.
19. R1-2402801 Discussion on other aspects of AI/ML model and data Continental Automotive
20. R1-2402844 Discussion on other aspects of AI/ML model and data InterDigital, Inc.
21. R1-2402850 Additional study on other aspects of AI model and data NVIDIA
22. R1-2402873 Discussion on other aspects of AI/ML model and data Apple
23. R1-2402922 On aspects of AI/ML model and data framework Lenovo
24. R1-2403000 Other Aspects of AI/ML Model and Data Nokia
25. R1-2403014 Discussion on other aspects of AI/ML model and data ETRI
26. R1-2403148 Other Aspects of AI/ML framework AT&T
27. R1-2403186 Other aspects of AI/ML model and data Qualcomm Incorporated
28. R1-2403236 Discussion on other aspects of AI/ML model and data NTT DOCOMO, INC.