**3GPP TSG RAN WG1 #117 R1-240XXXX**

**Fukuoka City, Fukuoka, Japan, May 20th – 24th, 2024**

Source: Moderator (OPPO)

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

Agenda Item: 9.1.3.3

Document for: Discussion and Decision

# Introduction

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

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

Accordingly, RAN1 chair arranged several agenda items for different topics, among which agenda item 9.1.3.3 focuses other aspects of AI/ML model and data including model identification/procedure, training data collection for UE-sided model, and model transfer/delivery. The corresponding objectives captured in the Rel-19 WID (RP-213599) is copied as below for reference:

|  |
| --- |
| Study objectives with corresponding checkpoints in RAN#105 (Sept ’24):* …
* Necessity and details of model Identification concept and procedure in the context of LCM [RAN2/RAN1]
* CN/OAM/OTT collection of UE-sided model training data [RAN2/RAN1]:
	+ For the FS\_NR\_AIML\_Air study use cases, identify the corresponding contents of UE data collection
	+ Analyse the UE data collection mechanisms identified during the FS\_NR\_AIML\_Air (TR 38.843 section 7.2.1.3.2) study along with the implications and limitations of each of the methods
* Model transfer/delivery [RAN2/RAN1]:
	+ Determine whether there is a need to consider standardised solutions for transferring/delivering AI/ML model(s) considering at least the solutions identified during the FS\_NR\_AIML\_Air study
 |

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

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

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

# Model identification/procedure

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 1: For MI-Option 1, conclude that an associated ID is a NW-specific ID, and the network assigns/manages associated IDs.**Proposal 2: For MI-Option 1, conclude that an associated ID is not a model ID. Remove Alt.3 in the agreement.** *~~Alt.3: Associated ID(s) is assumed as model ID(s)~~*
	+ *~~“Model ID is determined/assigned for each AI/ML model” in D is not needed.~~*

*Observation 1: For MI-Option 1, one associated ID may be mapped to multiple models trained using the data collected based on the associated ID.**Proposal 3: For MI-Option 1, conclude that RAN1 only discuss the case that model IDs are assigned in the procedure.* * *The procedure without Step D is not considered as model identification and it can be a separate discussion (e.g., for alignment of additional conditions).*

*Proposal 4: For MI-Option 1, conclude that model IDs are assigned only by the NW (Alt. 1).**Proposal 5: Clarify the following for MI-Option 2 before further discussion.** *The boundary between MI-Option 1 and MI-Option 2, 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 6: Study the following, if MI-Option 2 is supported.** *Method of referring to a dataset, e.g., whether we can use the associated ID in MI-Option 1.*
* *Necessity of dataset transfer and the mechanism of doing it.*

*Proposal 7: Support MI-Option 3 with further study of its procedures and other details, if the mechanism of model transfer is determined (either 3GPP transparent or non-3GPP-transparent).**Proposal 8: For MI-Option 4, if justified, clarify the relationship between the reference model and multiple derived models, in the case only the structure of the reference model is standardized.**Proposal 9: Proponents to justify MI-Option 5 as one of the valid options for model identification.**Proposal 10: For MI-Option 5, if justified, clarify whether MI-Option 5 (Model identification via model monitoring) requires performance monitoring of unidentified and inactive models.* |
| Ericsson[2] | *Observation 1 Information and/or indication on NW-side additional conditions and NW configuration by the NW can be considered as implicit model identification initiated by the NW.**Observation 2 The applicability signalling of a functionality in a reactive approach mitigates the need for model-ID based LCM.**Observation 3 RAN1 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 4 RAN1 has yet not concluded on the support of training collaboration type 1 and therefore, there is still uncertainty in the need to support Model identification in model transfer from NW to UE.**Proposal 1 Conclude that information and/or indication on NW-side additional conditions by the NW can be considered as implicit model identification initiated by the NW.**Proposal 2 For MI-Option 1, in step A, consider the following use-case information as the “data collection related configuration(s) and it/their associated ID(s)” which is transmitted from NW to UE,**a. Beam management: data collection related configuration(s) comprises the set A/B configuration, and the associated identifiers comprises consistency information of the NW transmission parameters when transmitting set A/B.**b. Positioning: data collection related configuration(s) and their associated IDs, including:**i. The validity area of the model for inference, e.g., a list of TRPs that transmits PRS**ii. TRP/ARP location information**iii. Spatial domain information of PRS**iv. Time synchronization information of the TRPs**v. PRS configuration information**Proposal 3 For MI-Option 1, AI-example1, conclude that step A/B/C is the baseline method. The need for step D is not justified.**Proposal 4 For MI-Option 1, add a note that “Associated ID(s) can be considered as a logical model ID(s)”, and how the UE maps the associated ID to a possible physical model ID is transparent to the NW.**Proposal 5 For MI-Option 1, further study its applicability to the two-sided use case.**Proposal 6 For Ml-Option 2,3, and 4, RAN1 to conclude that they are not applicable for the UE-sided model use cases.**Proposal 7 For Ml-Option 2,3, and 4, RAN1 to conclude that there is no need to discuss until further progress is made for the two-sided CSI compression use case.**Proposal 8 MI-Option 5 should not be considered further.* |
| Huawei[3] | *Observation 1: The boundary between model identification and functionality identification for the Functionality with model ID is not clear.**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.**Observation 3: The feasibility of MI-Option 1 for one-sided model is unclear, since introducing globally unique model ID for the purpose of data categorization indication is not really helpful to the UE side and may harm the proprietary preservation of the NW side.**Observation 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.**Proposal 1: Consider functionality-based identification/LCM with model ID as the same category with model-ID-based identification/LCM until further clarification on the difference is achieved.**Proposal 2: For studying the applicable sub use cases of model identification and model-ID-based LCM, take two-sided model as the starting point.**Proposal 3: MI-Option 1 for one-sided model with globally/area unique model ID is deprioritized.* *Proposal 4: For ensuring consistency between training and inference via data collection related configuration(s) and/or indication(s), study associated ID under functionality identification subject to cell specific manner as a starting point.**Proposal 5: MI-Option 2 is applicable to two-sided model case.**Proposal 6: For the transmitted information of MI-Option 2, if the dataset is delivered from NW side to UE side, the following information may be needed:** Input and output of the NW side CSI generation part for training the UE side CSI generation part.** Other meta information, including at least: dataset ID, size of dataset, type/format of data samples, model scalability information, quantization method for CSI feedback.**Proposal 7: For the procedure of MI-Option 2, the model identification is achieved when the dataset ID is delivered in together with the delivered dataset.**Proposal 8: MI-Option 3 is applicable to two-sided model case.**Proposal 9: For the transmitted information of MI-Option 3, taking Case z4 for example, the following information may be needed:** Model parameters.** Other meta information, including at least: model ID, format of the parameters, model structure information, quantization method and parameters.**Proposal 10: For the procedure of MI-Option 3, the model identification is achieved when the model ID is delivered in together with the delivered model.* |
| Intel[4] | *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 of dataset transfer and model transfer for MI-Options 2 and 3 respectively.**Proposal 2:**• For MI-Option 1, on determination/assignment of model ID(s), the following options are considered further:**o (Opt. A) Model(s) ID(s) are already determined/assigned prior to assignment of Associated ID.**o (Opt. B) Model(s) ID(s) are assigned/determined at the time of association to the configuration(s) and/or indication(s), i.e., following assignment of Associated ID.**Observation 3:**• [Relationship between model ID(s) and Associated ID(s) for Alt. 1/2/4] For MI-Option 1, if model ID(s) are already assigned/determined prior to assignment of Associated ID,**o the assigned/determined model IDs for the models reported by the UE are simply associated with the assigned Associated ID(s) without any inherent relationship between model ID(s) and the Associated ID;**o a single model, identified by a model ID, may be reported for and thus map to multiple Associated ID(s);**o it is possible that none of the identified models may be reported in response to assignment of an Associated ID for a given data collection configuration/indication.**Observation 4:**• [Relationship between model ID(s) and Associated ID(s) for Alt. 1/2/4] For MI-Option 1, if model ID(s) are assigned/determined at the time of association to the configuration(s) and/or indication(s), i.e., following assignment of Associated ID,**o model ID(s) for the reported model(s) can be determined/assigned to have a hierarchical relationship to an Associated ID, i.e., follow Associated ID;**o a single model may be identified using multiple model IDs if the model may be consistent with different Associated IDs for different data collection related configurations/indications.**Observation 5:**• For MI-Option 1, if Associated ID is assumed as model ID,**o multiple models may be associated with an Associated ID corresponding to a set of configuration(s) and/or indication(s) for data collection and share a common model ID;**o depending on the number of models reported for an Associated ID, this alternative may provide a level of control for LCM operations that lie on the continuum between functionality- and model-level LCM.**Proposal 3:**• MI-Option 1 can be applicable and beneficial for all the identified use-cases considered during Rel-19 (beam management, positioning, CSI prediction, and CSI compression) that would benefit from model-level granularity for LCM operations for a given functionality.**Proposal 4:**• For MI-Option 2, 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 5:**• 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.* |
| Spreadtrum[5] | *Observation 1: MI-Option2/3/4 can be considered for two-sided use cases**Proposal 5: For AI-Example of MI-Option 1, it is up to NW to assign the model ID.**Proposal 6: MI-Option 5 can be deprioritized.* *Proposal 7: For two-sided model, model identification, e.g., MI-Option2/3/4, can be considered, at least for the sake of providing pairing of two-sided models.* |
| IDC[6] | *Observation 1: In beam management use case, for the cases of AIML models only at network side, the LCM procedures can be network implementation specific, and the model identification may not be necessary.* *Observation 2: In positioning use case 3a, if LCM is performed by the LMF, NW-side model identification may be necessary since the LMF may need to know information about AIML model(s) implemented at the gNB.**Observation 3: Functionality-based LCM and model-ID-based LCM may be applicable for potentially different use cases, model deployments, model management granularity and collaboration levels.**Observation 4: Model-ID based signaling in a Functionality is beneficial for model-level management (e.g., pairing of models) of UE-part of two-sided models.* *Observation 5: based on how to assign Associated ID, associated ID level LCM may be equivalent to model-ID based LCM. Therefore, Alt.3 seems to be enough to cover most of use case for the single-sided model developed and used at the UE side.**Proposal 1: An associated ID corresponds to data collection related configuration(s) and NW-sided additional conditions.**Proposal 2: For single-sided model developed and used at the UE side, down-select Alt-3 only for MI-Option 1 for further study.* |
| Samsung[7] | *Proposal#1: To study the necessity of MI-Option1, RAN1 to consider its application on model-level management of AI/ML operations at the UE including* *• Timeline management for LCM operations, e.g., model inference, activation, switching* *• Network’s awareness on UE’s AI/ML processing units and corresponding occupancy**Proposal#2: RAN1 to conclude ensuring consistency on network-side additional condition between model training and inference does not necessitate model identification. Indication on network-side additional condition based on Step A/B/C of AI-Example1 and additional interaction of associated IDs between UE and NW is sufficient.* *Proposal#3: 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: 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#4: For MI-Option 1 and MI-Option 2, consider the following additional procedure for model-ID-based LCM with model identification Type B1**• For NW’s indication on NW-side additional condition: The network provides the list of indicator(s) of network-side additional conditions for an AI/ML-enabled feature/FG**• For UE’s model identification to the network: The UE identifies a model with information on the supported configurations/conditions for AI/ML-enabled feature/FG and/or associated indicators for NW-side additional conditions.**• For model-ID based LCM: Network use model ID(s) for the identified model(s) to give LCM assistance, e.g., model activation, inference, monitoring, deactivation.* *Proposal#5: For 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#6: 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.* |
| vivo[8] | *Observation 1: Associated ID and model ID have different underlying logic.** Associated ID represents certain NW-sided implementation/configurations and/or wireless channel environments.* * Model ID represents certain AI/ML model implementation, which may require additional control/awareness of model beyond associated ID.**Observation 2: Directly using data categorization information as model ID is not future-proof for cases where real model-level awareness is needed.* *Observation 3: Directly using associated ID as model ID is not future-proof for cases where real model-level awareness is needed.* *Observation 4: Model ID could achieve refined model level control and might be needed in the following cases.* * Model switching timeline alignment across two sides;* * Model selection with appropriate performance target and complexity tradeoff;** Model monitoring metric calculation;**Proposal 1: To address the issue of maintaining consistency between training and inference, associated ID can be used rather than model ID. Model ID would be used for two-sided model pairing and model transfer.**Observation 5: Global associated ID may expose deployment choices of NW side, but is useful information to maintain consistency between training and inference.**Observation 6: Local associated ID either requires huge or infeasible efforts at UE side to categorize the collected data or may require cell/site/region specific model development and management.* *Proposal 2: Global associated ID can be optionally supported.**Proposal 3: Local associated ID can be supported with the understanding that the model is managed in a cell/site/region specific way.**Observation 7: If model level awareness is justified in Rel-19, global model ID can be used to indicate certain globally unique AI/ML models.* * Pros: Global model ID may avoid reporting information of UE AI/ML model and corresponding associated ID(s) in every cell for every UE for previously identified model** Cons1: Specification effort of how global model ID is assigned between different vendors is needed. Core network may be involved to manage the global model ID.* * Cons2: Global model ID may potentially expose vendor/device type information of UE side.**Observation 8: If model level awareness is justified in Rel-19, local model ID can be used to work with the ID information reported per indicates certain cell/site/region.** Pros1: Specification effort of how local model ID is assigned between different vendors is not needed. Core network may be not involved to manage the local model ID.** Pros2: Local model ID may potentially expose less deployment choices of UE side and/or NW side.** Cons: Local model ID should be assigned after UE reporting associated IDs in model identification when UE re-connect to the network every-time.**Observation 9: 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 10: Feasibility of model identification with dataset transfer is dependent on the feasibility of dataset transfer itself.**Proposal 4: Model identification is needed for cases where multiple models are transferred from NW to UE.* *Proposal 5: Multiple models transferred to UE could reduce the latency of model switch of the transferred models.**Proposal 6: 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 7: Model identification via standardization of reference models may have the following procedures:** MI-Option 4-1: UE may report specified (global) model ID of reference model. Specified (global) model ID is used for model control and performance monitoring.** MI-Option 4-2: UE may report specified (global) model ID of reference model. Then NW assigns local model ID for specified (global) model ID. Local model ID is used for model control and performance monitoring.**Proposal 8: Reference model may be also used in one-sided case. For example, RAN4 may also define some reference model for one-sided case.**Proposal 9: MI-Option 4 (model identification via standardization of reference models) can be used in cases when multiple reference models are specified, which would have the following purpose/usage.** Would partially ensure consistency between training and inference, where multiple reference models are specified considering more additional conditions from vendors;** Can support different AI model with different capabilities, if multiple reference models with different capabilities are pre-defined.**Proposal 10: How MI-Option 5 (model identification via model monitoring) works is not clear.* |
| Apple[9] | *Proposal 1: In MI-option 1, the associated ID(s) is assumed as model ID(s).* *Proposal 2: 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.*  |
| 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: In AI-Example1, model ID is assumed to be independent from associated ID.**Proposal 3: In AI-Example1, the applicable range of associated ID is assumed to be:**- Per cell (baseline);**- Per cell group (can be considered);**- Other ranges (per NW vendor, per PLMN or global) need more justification and clarification on feasibility, difficulty, proprietary issue and offline coordination engineering effort.**Proposal 4: In AI-Example1, model ID is assigned by network after UE reporting the information of its AI/ML models to the network.**Proposal 5: In AI-Example1, meta information (if supported) of an AI/ML model carries all related associated ID(s) of the AI/ML model, and is transmitted from UE to network.**- Meta information may carry other information, which is up to future discussion.**- The procedure and signaling of meta information transmission is out of RAN1.* *Proposal 6: Unless clear additional benefit is found and justified for MI-Option1 compared to functionality-based LCM with associated ID, do not support MI-Option1.**Proposal 7: For MI-Option 2, the following example can be considered:**- AI-Example2** Step A, training data collection phase, NW indicates an associated ID(s) to UE/UE-side, associated with the transferred dataset(s), representing NW-side additional condition;** Step B, UE/UE-side receives dataset(s) corresponding to associated ID(s);** Step C, AI/ML models are developed (e.g., trained, updated) at UE side based on the received dataset(s) corresponding to the associated ID(s).;** Step D, inference phase, UE reports information of its AI/ML models corresponding to associated IDs to the NW. Model ID is determined/assigned for each AI/ML model. NW controls LCM of UE-sided model via model ID, based on reported associated ID(s) and current NW additional condition.**- Functionality-based LCM with associated ID, mirroring AI-Example2** Step A, training data collection phase, NW indicates an associated ID(s) to UE/UE-side, associated with the transferred dataset(s), representing NW-side additional condition;** Step B, UE/UE-side receives dataset(s) corresponding to associated ID(s);** Step C, AI/ML models are developed (e.g., trained, updated) at UE side based on the received dataset(s) corresponding to the associated ID(s);** Step D, inference phase, current NW indicates current associated ID(s) 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 associated ID(s) of training dataset(s) and current associated ID(s).**Proposal 8: For MI-Option 2, further study on dataset ID is needed, including:**- How dataset ID(s) is determined/assigned;**- Relationship between dataset ID(s), model ID(s) and the associated ID(s).**Proposal 9: 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.* |
| CMCC[11] | *Proposal 1: The following aspects could be the starting point when discussing the information of model during model identification:** The related functionality/AI enabled feature of model** Model’s applicable scenarios, configurations** Type/dimension of model input/output**Proposal 2: The following alternatives can be considered to resolve the NW-side additional consistency issue over multi cells or cell groups without Step D:**• Alt 1: Model transfer/delivery when UE need to move another cell or cell group**• Alt 2: NW-side additional conditions indication to UE side**• Alt 3: Offline inter-vendor collaboration, including gNB-gNB and/or gNB-UE collaboration**Proposal 3: It is suggested to deprioritize Alt 3 and Alt 4 for model ID(s) determination/assignment:**• Alt.3: Associated ID(s) is assumed as model ID(s)**• Alt.4: Model ID is determined by pre-defined rule(s) in the specification**Proposal 4: It is suggested to further study Alt 1 and Alt 2 for model ID(s) determination/assignment:**• Alt.1: NW assigns Model ID**• Alt.2: UE assigns/reports Model ID**Proposal 5: 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 6: For MI-Option 3, it may include the following procedure:** Step1: NW may transmit the owned or configurable model list to UE.** Step2: UE will report supported model list to the NW.** Step3: NW may transfer model and assign the model ID to UE side for the following model deployment, model inference and corresponding LCM operation.**Proposal 7: It is suggested to deprioritize MI-option 4 for model identification.**Observation 1: There are two different directions for UE-sided model(s) development:**• Direction 1: Step A+B+C** It mainly is used for the procedure of functionality based LCM**• Direction 2: Step A+B+C+D** It mainly is used for the procedure of model-ID based LCM**Observation 2: If only associated ID is indicated to UE-side, and UE-side does not know the detailed meaning of NW-side additional conditions, then the consistency of NW-side additional conditions can only be maintained per cell or cell group.**Observation 3: One model could correspond to multiple associated IDs, or one associated ID could correspond to multiple associated IDs.* |
| Lenovo[12] | *Proposal 1: MI-Option 4 can be further discussed with inter-vendor training collaboration Option 1 (fully standardized reference model (structure + parameters)).**Proposal 2: De-prioritize the discussion on MI-Option 5.**Proposal 3: Further study the model identification procedures of the three options for the UE-part of two-sided models, e.g., to assist model pairing, in the AI/ML-based CSI compression use case.**Proposal 4: Define a set of data collection configuration(s) with associated ID(s) to 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: During both phases of data collection for training and inference, support procedures/signaling enabling UE and NW to exchange information related to data collection configuration(s) with the associated ID(s).**Proposal 6: Study MI-Option 2 with inter-vendor training collaboration Option 4 (Standardized data / dataset format + Dataset exchange between NW-side and UE-side) in the AI/ML-based CSI compression use case to ensure consistency.**Proposal 7: Study the association between the data collection related configuration(s) and the dataset with an association ID for model identification.**Proposal 8: The following procedure as an example of MI-Option 2 can be considered for further study (including the feasibility/necessity):**• Step 1: NW transfers the dataset(s) with it/their associated ID(s)* *o Note: The ID(s) can be also associated with the data collection configuration(s).**• Step 2: UE(s) receives the dataset(s) corresponding to the associated ID(s)* *• Step 3: AI/ML models are developed (e.g., trained, updated) at UE side based on the received dataset(s) corresponding to the associated ID(s).* *Step 4: UE indicates to the NW the accomplishment on the AI/ML model(s) development corresponding to the associated IDs.**FFS: the content and transfer of a dataset over the air interface.**Proposal 9: Study MI-Option 3 with inter-vendor training collaboration Option 3 (Standardized reference model structure + Parameter exchange between NW-side and UE-side) and Option 5 (Standardized model format + Reference model exchange between NW-side and UE-side) in the AI/ML-based CSI compression use case.**Proposal 10: The following procedure as an example of MI-Option 3 can be considered for further study (including the feasibility/necessity):**• Step 1: NW transfers the model(s) with it/their associated ID(s)**o Note: The ID(s) can be associated the data collection configuration(s).**• Step 2: UE(s) receives the model(s) corresponding to the associated ID(s).**• Step 3: AI/ML models are developed (e.g., trained, updated) at UE side corresponding to the associated ID(s).* *• Step 4: UE reports the associated IDs of the developed model(s) to the NW.**FFS: the content and transfer of a model over the air interface* |
| NVIDIA[13] | *Proposal 1: Conclude that there is a need for model identification in the context of LCM.**Proposal 2: Study 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 4: Model identification via standardization of reference models. (for CSI compression)**• MI-Option 5: Model identification via model monitoring* |
| LGE[14] | *Proposal#3. Clarify that any LCM that does not require assigning model ID belongs to functionality-based LCM.**Proposal#4. In AI-Example1, Step A/B/C are sufficient for UE-sided model to address the consistency issue between training and inference.**Proposal#5. For AI-Example1, Step D may be useful for two-sided model for model pairing but not for UE-sided model. In addition, the Alt3 of Step D in the examples needs further clarification on how to ensure one-to-one mapping between associated ID and each UE-sided model.**Proposal#6. 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.* |
| Fujitsu[15] | *MI-Option1**Observation-1: For MI-Option1, if the associated ID is assumed as a local ID, the following problems can be observed:** For the same associated ID, its corresponding NW-side additional conditions across cells may be different and cause data feature ambiguity in the data categorization for model training.* * The data feature ambiguity in model training may have impacts on model generalization capability.**Observation-2: For MI-Option1, if the associated ID is assumed a global ID, a common mapping rule between the ID(s) and the NW additional condition(s) may need to be specified. If so, the following problems can be observed:** The potential restrictions on NW implementation.** The potential risk of disclosing NW vendor’s proprietary information.* * The workload on collecting dataset with sufficient global IDs for model training.**Observation-3: In MI-Option1, the same associated ID across training-inference may only indicate the similarity of NW-side additional conditions, and cannot be used as the proof to predict that the UE-side model is workable.**Observation-4: In MI-Option1, the different associated IDs across training-inference may only indicate the difference of NW-side additional conditions across training and inference. They cannot differentiate the data features at UE side and be used as the proof to predict that the UE-side model is not workable.**Proposal-1: For MI-Option1, UE’s assumptions on the same associated ID needs to be clarified:** Alt-1: same NW additional conditions across training and inference within a cell.** Alt-2: same NW additional conditions across cells.**Proposal-2: For MI-Option1, UE’s assumptions on its model’s applicability and decision on model activation cannot rely on the associated ID(s) alignment across model training stage and model inference stage.**MI-Option5**Observation-5: For MI-Option5, the UE-side model is identified by the NW from model’s applicability perspective.* *Observation-6: For MI-Option5, the model selection procedure via model monitoring is assumed to be done at UE-side.* *Observation-7: For MI-Option5, the NW-side additional conditions during model monitoring/selection procedure are assumed as unchanged.**Proposal-3: For MI-Option5, the model ID assigned from NW is for identifying the model’s applicability under certain NW-side additional conditions. It can be named as model applicable ID and is assumed as a local ID.**Proposal-4: Regarding the model applicable ID assigned from NW to UE in MI-Option5, the following aspects are suggested for further study:** The association between the ID and its corresponding NW-side additional conditions.** The linkage between the model applicable ID and the associated ID when the associated ID in MI-Option1 is available to MI-Option5.** The relationship between model ID and the model applicable ID.**Proposal-5: The procedures of MI-Option5 are further clarified as below:** UE initiates the model monitoring/selection procedure, and NW provides necessary measurement configurations accordingly.** Applicable model(s) is selected via model monitoring under certain NW-side additional conditions.** NW assigns model applicable ID(s) to the selected model(s)** UE reports its model applicable ID(s) of a cell to the NW when it is available at the UE. The NW can decide activation of the corresponding model and skip the model monitoring/selection.**Proposal-6: The following approaches are suggested to be studied as the way to reduce the monitoring cost in MI-Option5:** The number of model candidates for monitoring can be controlled, e.g. by configuration alignment, by the associated ID.** The model monitoring procedure can be skipped if the model applicable ID for a cell is reported from the UE.** Model input-based monitoring can be considered for model selection.** Monitoring termination conditions can be provided to UE to control the latency and power consumption in model monitoring.**Proposal-7: The combination of MI-Option1 and MI-Option5 is suggested to be studied to ensure performance-level consistency across training-inference.* |
| Xiaomi[16] | *Observation 6: Compared with approach of step A/B/C and additional interaction of associated IDs between UE and NW, MI-Option 1 is still beneficial considering the following aspects* * Potential processing interruption management** Reducing network burden in handling the additional condition**Observation 7: MI-Option 1 is applicable to one-sided model**Proposal 3: The associated ID is not equivalents to the model ID**Proposal 4: 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* *- Step 1: NW transfers the dataset to UE and dataset ID is assigned.* *- 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.**- Step 3: UE side reports the existence of the model together with the associated with dataset ID and other meta information**- Step 4: NW assigns the model ID to the model**- Step 5: UE reports the model ID to indicate the availability of the model**Proposal 5: Consider the following procedure for MI-Option 3** Step 1: model identification from NW to UE, meta information and model ID would be shared* * Step 2: UE confirms the model transfer or delivery* * Step 3: Model transfer/delivery from NW to UE* * Step 4: UE reports the model ID to indicate the availability of the model**Proposal 6 : Consider the following procedure for Type A model identification** Step 1: Data set construction* *- Option 1: Dataset is obtained via offline coordination**- Option 2: Via data collection from UE* * Step 2:**- Train/Update the AI model offline** Step 3:**- UE side reports the model information offline. The reported information may include model input, output, associated network additional condition, performance and potential processing time for model activation or switch**- NW side assigns the model ID for this model to UE side offline** Step 4:**- UE reports the model ID to network to indicate the availability of the model**Proposal 7: Associated ID can be considered for data collection for type A model identification* |
| NEC[17] | *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.**Observation 2: The associated ID(s) is used for at least ensuring the consistency across training and inference and one or more associated ID(s) can be attached to one same model ID to reflect different NW side additional conditions.**Observation 3: 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 1: Support model ID and model identification in Rel-19.**Proposal 2: RAN1 should study following options for model identification Type B for further discussion.**− MI-Option 1: Model identification with data collection related configuration(s) and/or indication(s)**− MI-Option 3: Model identification in model transfer from NW to UE**Proposal 3: It is necessary to clarify the definition of associated ID for each use case separately.**Proposal 4: For MI-Option 1, further clarify the data collection related configurations for each use case. And support UE to request the needed data collection related configurations.**Proposal 5: In the model identification procedure, support the combination of Alt.1: NW assigns Model ID and Alt.2: UE assigns/reports Model ID, which is used to link a UE reported global model ID to a NW assigned local model ID.**Proposal 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* |
| Google[18] | *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.* |
| ZTE[19] | *Proposal 1: In Rel-19 AI/ML framework study, type B model identification is prioritized compared with type A model identification.* *Proposal 2: 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 3: In Rel-19 AI/ML framework study, in order to support a complete and unified solution for model identification, multi-vendor collaboration, and model pairing, MI-Option 2, MI-Option 3, and MI-Option 4 are prioritized.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Model identification | Multi-vendor collaboration | Whether model pairing is addressed | Analysis |
| Data collection related configuration(s) and/or indication(s) | MI-Option 1 |  | No | MI-Option 1 may not address the model pairing issue directly. For example, even with the same data collection related configurations and/or indications, the UE-part model may not be compatible with the network-part model due to the quantization method, data resolution, output size of the CSI generation part, etc. |
| Dataset | MI-Option 2 | Option 2, Option 4 | Yes | MI-Option 2 (dataset transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue. |
| Model transfer | MI-Option 3 | Option 3, Option 5 | Yes | MI-Option 3 (model transfer) can be applied to address the multi-vendor collaboration issue and model pairing issue. |
| Standardization of reference models | MI-Option 4 | Option 1 | Yes | MI-Option 4 (standardized reference model) can be applied to address the multi-vendor collaboration issue and model pairing issue. |
| Model monitoring | MI-Option 5 |  | No | MI-Option 5 (model monitoring) may not address the model pairing issue directly. For example, in case of poor model performance, it is not clear whether it is due to the incompatibility of the model or other potential reasons, e.g., additional condition.  |

*Observation 1: Comparison of three flavours of “MI-Option1”is as following.*

|  |  |  |
| --- | --- | --- |
| Flavour | Detailed steps | Analysis |
| Flavour#1 | Step A + Step B + Step C + Step D1Step D1: UE reports the Functionality information to the base station based on the associated ID indicated by the base station and availability of UE’s model.  | UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.Based station performs Functionality-based LCM based on the Functionality information reported by the UE |
| Flavour#2 | Step A + Step B + Step C + Step D2Step D2: UE reports the Functionality information and associated ID supported by the UE to the base station based on the availability of UE’s model. | UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.Based station performs Functionality-based LCM based on the Functionality information reported by the UEThe difference between Flavour#1 and Flavour#2 is that, in addition to the Functionality information, UE also indicates the supported associated ID to the base station with Flavour#2. The advantage of Flavour#2 is that, once base station updates the associated ID, UE doesn’t need to report the supported Functionality again. Base station can determine the supported Functionality of the UE by the reported associated ID from UE.  |
| Flavour#3 | Step A + Step B + Step C + Step D3Step D3: UE reports the Functionality information and associated ID supported by the UE to the base station based on the availability of UE’s model. Model ID is assigned by the base station. Model-ID based LCM is performed.  | UE is responsible to determine the supported Functionalities based on the associated ID indicated by the base station and availability of UE’s model.After assigning the model-ID, based station performs model-ID based LCM. |

*Proposal 4: If model ID is needed for MI-Option1, model ID is assigned by the base station instead of UE.* *Observation 2: MI-Option 1 can’t address the model pairing issue for two-sided model.* *Observation 3: 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 4: 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 5: 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 6: 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.* |
| Panasonic[20] | *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.**Proposal 1: Before model development in UE side, UE needs to transfer the collected data. Whether over the top or through NW should be concluded in RAN2. No difference would be seen from RAN1 side.**Proposal 2: The store location of the developed model is within NW or outside of NW (i.e. case y and case z1) should be concluded in RAN2 as no difference would be seen from RAN1 side.* *Proposal 3: The information of its AI/ML models corresponding to associated IDs can be the model is available "within UE side" or available "within the UE". These can be discussed for use case specific with the consideration of the required latency and the number of models.* |
| MediaTek[21] | *Observation 2-1: Ensuring the consistency could be done with model identification**Proposal 2-1: The model identification for training data collection could be supported for the case that, when a model is identified with a model ID, UE is not expected to have any further change on the NW side additional conditions**Proposal 2-2: For step D, it may also consider the condition that the model ID may not be assigned to abort the model development**Proposal 2-3: Consider NW to assign model ID in step D**Proposal 2-4: Associated ID(s) could be discussed in the respective use case agenda item* |
| ETRI[22] | *Observation 1: The Associated ID can be used to configure and manage datasets generated through the data collection process.**Observation 2: By utilizing the Associated ID, the NW can manage the dataset transfer and related information.**Proposal 1: Model-ID-based LCM can be integrated with Functionality-based LCM by using model ID for LCM operations.**Proposal 2: Associated IDs are configured and managed for each functionality.**Proposal 3: 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.**Proposal 3: Association ID and Model ID can have a many-to-many relationship.**Proposal 4: From RAN1 perspective, for UE-sided model(s) developed at UE side, following procedure is an example (noted as AI-Example2) of MI-Option2 for further study (including the feasibility/necessity).**A: For data transfer, the NW signals the dataset related information(s) and its/their associated ID(s).* *B: UE(s) receive the dataset(s) corresponding to the associated ID(s).**C: AI/ML models are developed at UE-sided based on the received dataset(s) corresponding to the associated ID(s).* *D: UE reports information of its AI/ML models corresponding to associated IDs to the NW. Model ID is determined/assigned for each AI/ML model.* |
| OPPO[23] | *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 B MI-Option 1,**• Step D should be supported for the UE not involved in Step A, B and C.**o Alt.1: NW assigns Model ID is preferred because it supports model identification for UE involved or not involved in Step A, B and C.**o Alt.2 is not preferred unless advantage over Alt.1 can be justified.**o Alt.3 is not preferred because it only supports model identification for UE involved in Step A, B and C.**o Details needs to be clarified for Alt.4.**• Strive for achieving the 1-to-1 mapping between model ID(s) and the associated ID(s), thus for the same inference behavior for UE involved or not involved in Step A, B and C.**Proposal 3:* *For model identification type 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).**Proposal 4:* *For other options of model identification type B,**• 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.**• FFS the name of the ID (e.g. model ID, dataset ID, additional condition ID).**Proposal 5: Functionality ID can be used for indication functionality between NW and UE.**Proposal 6: At least for LCM with non-3GPP-based model transfer, Local model ID can be a simple number, which is similar to the resource/configuration ID in the legacy NR specification and does not include explicit information about the model, e.g., scenarios/configurations/sites.**Proposal 7: 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 8: 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.* |
| Nokia[24] | *Proposal 1: For MI-Option 1, considering steps A -to - D, the following aspects are further applicable,* *• For associated ID,* *o Defining of associated IDs is up to the NW vendor implementations and shall not disclose any proprietary NW information.**o For BM use-cases, associated ID can be linked to CSI resource configuration (CSI-resourceConfig), or resource sets defined by a CSI-resourceConfig.* *o For positioning use-cases, associated ID can be linked to the PRS resource configuration (NR-DL-PRS-Info) or PRS resource sets (nr-DL-PRS-ResourceSet) defined by a NR-DL-PRS-Info.* *o In addition to the associated IDs, data collection configuration(s) may also associate with the global cell identities (GCIs).* *o A fixed bit field, e.g., 8/10 bits (provides max 256/1024 IDs) can be considered for the associated ID.**• For model-ID,* *o In Step D, the UE assigns model ID, and reports associated IDs (and optionally GCIs) related to the assigned model ID.* *• Model ID may be related to one or more associated IDs.* *o Reporting of model-IDs does not have to be in the UE-capability report. RAN1/2 to investigate further exact reporting details.*  |
| AT&T[25] | *Proposal 2: For Rel-19, support a unified LCM providing both functionality-based and model-ID-based operations.* *• Functionality-based operation is supported by default.**• Model-ID, if needed, can be used in the unified LCM for model ID based LCM operations.**Proposal 3: For both model identification Type B1 and B2* *• Network assigns the model ID(s) for the identified model(s) if model ID(s) assignment is needed**• FFS: How to define a model ID for assignment**Proposal 4:**• RAN1 focus on the study of MI-Option 1 (including the necessity) for the use cases of AI-based positioning, AI-based beam management.**• RAN1 focus on the study of MI-Option 1 (including the necessity) for the use case of CSI prediction.**• RAN1 focus on the study of MI-Option 2/3/4 (including the necessity) for the use case of CSI compression.**Proposal 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.* |
| DOCOMO[26] | *Observation 1: For the support of scenario/site specific models, the following aspects should be considered.**・(Training phase) How to prepare scenario/site specific models. In other words, how to prepare models specific to additional condition.**・(Inference phase) How to select an appropriate scenario/site specific model among prepared models. In other words, how to ensure consistency between NW side additional conditions and UE side model.* *Observation 2: Model identification changes management granularity from associated ID to model ID, which increases NW management burden and NW awareness of UE side performance.**Observation 3: MI-Option2 is applicable with two-sided model and one-sided model, where the procedure of MI-Option2 can be described as follows:**Step1: NW side obtains the information about supportable model at UE device.**Step2: AI/ML models are developed and stored at NW side.**Step3: NW transfers model the developed model with model ID.**Observation 4: MI-Optio3 is applicable with two-sided model and one-sided model, where the procedure of MI-Option3 can be described as follows:**Step1: NW transfers dataset with associated ID for certain functionality.**Step2: AI/ML models are developed at UE side based on the collected data.* *Step3: UE reports information of its AI/ML models corresponding to associated IDs.**Proposal 1: 3GPP should consider the framework to support scenario/site specific model.**Proposal 2: Future compatibility with model transfer and model storage at NW side should be taken into consideration after they are supported in 3GPP.* |
| Qualcomm[27] | *Observation 1: Alt. 3 of AI-Example1 which is representative of NW-side initiated model identification, can serve the purpose of helping with ensuring consistency of NW-side additional conditions across training and inference, and is a more reasonable alternative compared to other alternatives which may lead to increased NW-side LCM complexity.**Proposal 1: Regarding “how model ID(s) is determined/assigned” in Step D of AI-Example1 of MI-Option 1, support Alt. 3 in which associated ID(s) is assumed as model ID(s).**Proposal 2: 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 3: From RAN1 perspective, for UE-sided model(s) developed at UE side, following procedure is an example (noted as AI-Example2) of MI-Option1 for model identification Type A:**• A: For data collection, the data collection related configuration(s) and it/their associated ID(s) is coordinated offline between NW and UE**o Associated IDs for each sub use case in relation with NW-sided additional conditions**• B: UE(s) collects the data corresponding to the associated ID(s)* *• C: AI/ML models are developed at UE side based on the collected data corresponding to the associated ID(s).* *• D: The associated ID(s) in Step A is assumed as model ID(s), and UE reports supported AI/ML model IDs.**Note: Step A/B/C and additional interaction of associated IDs between UE and NW can be considered as a different solution for resolving the consistency without model identification.**Proposal 4: Support AI-Example2 of MI-Option1 for model identification Type A.**Proposal 5: From RAN1 perspective, for UE-sided model(s) developed at UE side, following procedure is an example (noted as AI-Example3) of MI-Option1 for model identification Type A:**• A: For data collection, NW signals the data collection related configuration(s) and it/their associated ID(s)* *o Associated IDs for each sub use case in relation with NW-sided additional conditions**• B: UE(s) collects the data corresponding to the associated ID(s)* *• C: AI/ML models are developed at UE side based on the collected data corresponding to the associated ID(s).* *• D: UE-side initiates model identification procedure in an offline manner and gets a model ID(s) assigned, and UE reports supported AI/ML model IDs.* *Proposal 6: From RAN1 perspective, for UE-sided model(s) developed at UE side, following procedure is an example (noted as AI-Example4) of MI-Option1 for model identification Type A:**• A: For data collection, the data collection related configuration(s) and it/their associated ID(s) is coordinated offline between NW and UE**o Associated IDs for each sub use case in relation with NW-sided additional conditions**• B: UE(s) collects the data corresponding to the associated ID(s)* *• C: AI/ML models are developed at UE side based on the collected data corresponding to the associated ID(s).* *• D: UE-side initiates model identification procedure in an offline manner and gets a model ID(s) assigned, and UE reports supported AI/ML model IDs.* |
| Continental Automotive[28] | *Proposal 1: Model ID is applied as basis for model identification related issues.**Proposal 2: MI-Option 1/2/3 can be prioritized and MI-Option 5 can be deprioritized.**Proposal 3: Associated IDs can be based on varying combinations of mapping relation indication depending on model operation use cases and/or LCM phases.**Proposal 4: Mapping relation information can be configured for model versions in association with additional condition grouping or segmentation related to model training.**Proposal 5: Paired additional conditions on both NW and UE sides can be identified for alignment with the pre-configured information (e.g., index or ID).* |
| IIT[29] | *Observation#1 : Regarding model delivery/transfer following tables shows the current status for various options for UE-sided and Two-sided model and relation with model storage locations and transfer/delivery type.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Case** | **Model delivery/transfer** | **Model storage location** | **Training location** | **UE-sided Model** | **Two-sided Model** |
| **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 | Deprioritized for Rel-19 |  |
| **z3** | model transfer in open format. | 3GPP Network | UE-side / neutral site | Deprioritized for Rel-19 | Deprioritized for Rel-19 |
| **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 | Deprioritized for Rel-19 | Deprioritized for Rel-19 |

*Proposal 4: In case of MI option -1, with option D, study feasibility of ALT3 of using associated ID(s) as model ID(s) atleast for enabling functionality based LCM**Proposal 5: Study in feasibility and details of MI-Option3 for Model transfer/delivery case Z4 in Rel 19* |

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

## 1st round discussion

#### **Proposal 2.1.1**

 For AI-Example 1 of MI-Option1, there are many tdocs discussing the associated ID(s). One discussion point is whether it is global ID or local ID. Based on the tdocs, it seems most companies at least can accept local ID. Thus, the following proposal is suggested for discussion.

**Proposal 2.1.1:**

**Regarding AI-Example1 of MI-Option1, the associated ID at least can be local ID**

* **Global cell identity (GCI) can be used together with the associated ID**
* **FFS: whether the associated ID can be global ID**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | We support the associated ID at least can be local ID. To use global cell identity (GCI) together means the ID is cell specific. There can be the operation that associated ID is not always cell specific in the following type of the scenario.- the association intended to cover multiple of cell relations- the association that can be same among multiple of cells ( for example, small cells of the same configuration are deployed in the same area)Even such case, we think it can be still local within a PLMN (i.e. operator). Therefore, we propose to modify as following.* **Global cell identity (GCI) or PLMN ID can be used together with the associated ID**
 |
| Huawei, HiSilicon | 1) To draw conclusion on the necessity of MI-Option 1, we need to analyse MI-Option 1 and the alternative solution of “Step A/B/C and additional interaction of associated IDs between UE and NW”. Therefore, the proposal should not only focus on MI-Option 1, but be fairly applicable to both solutions.2) The meaning of local ID is not so clear, so it is updated to “cell specific ID”3) CGI is already available to UE in SIB, so UE can use it by implementation if needed. So, a note is added here.**Regarding ~~AI-Example1 of MI-Option1,~~ the associated ID, it at least can be assigned as cell specific ID ~~local ID~~** * **Note: Global cell identity (GCI) can be used together with the associated ID by UE implementation**
* **FFS: whether the associated ID can be global ID**
 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.2**

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

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

**Proposal 2.1.2:**

**Regarding the relationship between model ID(s) and the associated ID(s) in AI-Example1 of MI-Option1, further study the following options (including the necessity/benefit)**

* **ID-Rel-Option1: One model ID is linked to one associated ID by one-to-one mapping**
* **ID-Rel-Option2: One model ID can be linked to multiple associated IDs and each associated ID is only be linked to one model ID**
* **ID-Rel-Option3: One associated ID(s) can be linked to multiple model IDs and each model ID is only linked to one associated ID**
* **ID-Rel-Option4: Model ID(s) can be linked to associated ID(s) by many-to-many mapping**
* **Note: Proponents of each option are encouraged to provide detailed analysis on applicable use cases, benefit, necessity and so on**

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



Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | We agree to list four options. Our view is all option are valid and its usage depends on what level of the model needs to be identified by NW. In ID-Rel-Option1, model ID is logical model. NW is not required to identify what physical model is actually used by UE in this case. The physical model usage is up to UE side.In ID-Rel-Option2, model ID is physical model. For testing (or conformance test) and evaluation purpose, NW may be required to identify the physical model. The physical model can be used for multiple associated ID. This option is used for such situation.In ID-Rel-Option3, model ID is physical model. For testing (or conformance test) and evaluation purpose, NW may be required to identify the physical model. Because of the version up of the same model or UE choose different model depending on UE judgement of the situation, one associated ID can be mapped to multiple of physical model. This option is used for such situation.In ID-Rel-Option3, model ID is physical model. This is the case of both ID-Rel-Option2 and ID-Rel-Option3 are used. For physical model case identification case, this is more generic. |
| Huawei, HiSilicon | Why do we need to discuss the options of mapping one/multiple associated ID to one/multiple model IDs? The options are only valid after we have confirmed the MI-Option 1 is necessary to one-sided model. These options will not impact the comparison between MI-Option 1 and the alternative solution of “Step A/B/C and additional interaction of associated IDs between UE and NW”.Therefore, **we can defer this proposal after the necessity of MI-Option 1 is justified**. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.3**

In the submitted tdocs, some companies discussed the boundary of MI-Option1 and MI-Option2. Meanwhile, some companies think it is not clear how MI-Option2 is working. Thus, it would be beneficial to make some good clarification/concrete examples to facilitate further discussion on MI-Option2. Correspondingly, a proposal is suggested with the following considerations:

* Similar to MI-Option1, we provide some concrete example(s) to facilitate further discussion
* As there are discussions for two-sided model in CSI compression session, this proposal focused on UE-sided model
* There may be some proposal on MI-Option2 for two-sided model later in this meeting or the next meeting. We can wait for more progress in CSI compression session for now.

**Proposal 2.1.3**

**From RAN1 perspective, for UE-sided model(s) developed (e.g., trained, updated) at UE side, following procedure is an example (noted as AI-Example2) of MI-Option2 for further study (including the feasibility/necessity)**

* **A: The dataset(s) along with its associated dataset ID(s) is transferred from the NW-side to UE or UE-side via 3GPP standardized data / dataset format**
	+ **Note: the dataset ID(s) is determined by NW-side**
* **B: AI/ML models are developed (e.g., trained, updated) at UE side based on the above dataset(s).**
* **C: UE reports information of its AI/ML model(s) corresponding to the above dataset(s) to the NW. Model ID is determined/assigned for each AI/ML model**
	+ **Relationship between model ID and associated dataset ID**
	+ **How model ID(s) is determined/assigned, e.g.,**
		- **Alt.1: NW assigns Model ID**
		- **Alt.2: UE assigns/reports Model ID**
		- **Alt.3: The associated dataset ID is used as model ID**
	+ **FFS: how to report**
	+ **Note: C is to facilitate AI/ML model inference**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | We support to discuss this direction. In MI-Option1, UE side additional condition is managed by UE side. Therefore, NW is not required to identify them. On the other hand, in above MI-Option2, what UE side additional condition is used to obtain/accumulate the dataset(s) in step A is matter. In order to have the consistency between "UE side additional condition of the dataset(s) in step A" and "UE side additional condition at the time of inference", either of the following would be required. 1) Dataset(s) in step A is obtained by each UE side. In this case, NW effort can be too large compared with MI-Option1. 2) UE side additional condition is standardized. We are not so sure this option is feasible or not. |
| Huawei, HiSilicon | 1) We do not see a significant difference between one sided model and two sided model for the procedure of MI-Option 2, so the proposal is updated to be common to one-sided and two-sided – as 9.1.3.2 has not discussed this issue, we can first have some conclusion at 9.1.3.3 to accelerate the whole progress, and reuse to 9.1.3.2; there is no duplicated discussion anyway.2) The assignment of model ID may not be needed, since the dataset ID can represent the model during the inference phase.**From RAN1 perspective, for UE-sided/UE part model(s) developed (e.g., trained, updated) at UE side, following procedure is an example (noted as AI-Example2) of MI-Option2 for further study (including the feasibility/necessity)*** **A: The dataset(s) along with its associated dataset ID(s) is transferred from the NW-side to UE or UE-side via 3GPP standardized data / dataset format**
	+ **Note: the dataset ID(s) is determined by NW-side**
* **B: AI/ML models are developed (e.g., trained, updated) at UE side based on the above dataset(s).**
* **C: UE reports information of its AI/ML model(s) corresponding to the above dataset(s) to the NW. ~~Model ID is determined/assigned for each AI/ML model~~**
	+ **Relationship between model ID (if needed) and associated dataset ID**
	+ **How model ID(s) (if needed) is determined/assigned, e.g.,**
		- **Alt.1: NW assigns Model ID**
		- **Alt.2: UE assigns/reports Model ID**
		- **Alt.3: The associated dataset ID is used as model ID**
	+ **FFS: how to report**
	+ **Note: C is to facilitate AI/ML model inference**
 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.4**

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

**Proposal 2.1.4**

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | Monitoring based solution is discussed in BM session, where it is for functionality-based LCM. In this session, we focus on the control with finer granularity (i.e., model level). Thus, the discussions related to performance/model monitoring in these two sessions are separate.  |
| Panasonic | Support |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.5**

Based on the tdocs, some companies think that even if IM-Option4 is supported, the UE and NW can achieve the same understanding on the AI model(s). For this case, UE only needs to report its supported AI model(s), which can be done by UE capability signaling. Thus, the following proposal is suggested for further discussion

**Proposal 2.1.5**

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

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | We are not so sure the meaning of "not pursued for Rel-19 normative work". Pursued or not can be discussed in RAN1 CSI compression session and RAN4 as you described. Important would be following. To ensure consistency between training and inference regarding NW-side additional conditions is feasible in ~~The model identification procedure dedicated to~~ IM-Option4 ~~is not pursued for Rel-19 normative work~~ |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.6**

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

**Proposal 2.1.6**

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Panasonic | Support |
| Huawei, HiSilicon | This proposal is contradictory with Proposal 2.1.3? Maybe it can also include MI-Option 3.**The model identification procedure dedicated to IM-Option2/MI-Option 3 for one-sided model is not pursued for Rel-19 normative work** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 2.1.7 (Placeholder)**

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

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

**Proposal 2.1.7 (proposal may be provide later)**

Companies can provide comments/inputs in the following table:

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

# Training data collection for UE-sided model

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Ericsson[2] | *Proposal 9 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.* |
| Huawei[3] | *Proposal 12: For the continued study of data collection for UE-side model training, lower the priority of the discussion at RAN1 due to the following reasons:** The content for use cases have already been provided in the Rel-18 LS reply from RAN1.** Discussion of UE data collection mechanisms is out of RAN1 scope.* |
| Intel[4] | *Proposal 8:**• 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.* |
| Spreadtrum[5] | *Proposal 1: For data collection for UE-side model training, support 1a or we could wait the progress of RAN2.* |
| IDC[6] | *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 AIML positioning purpose, support MI-Option 1 for model identification type B.**Proposal 4: 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 5: 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 6: Quality indicators for the CSI prediction model training dataset could include at least the RSRP and TDCP.* *Proposal 7: A ground truth label quality indicator is associated with a UE or PRU location**Proposal 8: Support both hard (1 or 0) and soft indicator (0, 0.1, 0.2, …, 1.0) for a ground truth label quality indicator**Proposal 9: 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 10: 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 11: The LMF is the only entity that can generate a ground truth label quality indicator associated with location information of UE or PRU**Proposal 12: 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 13: 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 14: 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.* |
| Samsung[7] | *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#7: 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.*  |
| Apple[9] | *Proposal 3: From RAN1 perspective, option 1-1a is sufficient and no additional requirement is identified to enhance data collection for UE side model training.*  |
| CATT[10] | *Observation 9: 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 11: 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.* |
| CMCC[11] | *Proposal 10: Regarding the UE side data collection mechanism, RAN2 could take the Reply LS on Data Collection Requirements and Assumptions (R1-2310681) as the baseline.* |
| NVIDIA[13] | *Proposal 3: Conclude that there is a need for collection of UE-sided model training data.* |
| LGE[14] | *Proposal#7. 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.* |
| Fujitsu[15] | *Proposal-8: 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.**Proposal-9: For potential additional conditions in data content, we suggest that:**• Both NW-side additional conditions and UE-side additional conditions for data categorization and training-inference consistency should be studied.**• Which aspects/details can be considered as additional condition 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-10: The quantization of data samples in data collection needs to be studied.* |
| Xiaomi[16] | *Proposal 8: The data content and related information included in RAN1 LS (R1-2310681) to RAN2 can be set as baseline* |
| Google[18] | *Proposal 4: Support the NW and UE to maintain the same understanding on when the UE can perform data collection.**Proposal 5: One associated ID can be mapped to one or multiple model ID(s)**• For UE-side model, the NW only configures the associated ID**• For two-side model, the NW and UE should maintain the same understanding on the model ID for model inference**o Model ID can be either configured by the NW or reported by the UE* |
| ZTE[19] | *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.* |
| ETRI[22] | *Proposal 5: Datasets should be categorized based on NW configurations and configured functionalities during the data collection process.**Proposal 6: The NW can request UEs to transfer collected data immediately for the purpose of categorizing the dataset.**Proposal 7: The UE needs a mechanism to categorize data samples according to changes in its settings.* |
| OPPO[23] | *Proposal 9: 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.* |
| DOCOMO[26] | *Proposal 3: When considering data collection toward OTT/UE side server, the ownership of data should be clarified first.* |
| Qualcomm[27] | *Proposal 7: The RAN1/RAN2 discussion should be focused on data collection for model training on the UE side, considering the following**- Direct transfer of the collected data to the server for data collection for UE side training (in a 3GPP transparent or 3GPP non-transparent method)**- Transfer of the collected data to the server for data collection for UE side training (via CN or OAM).**Observation 2: The actual input/output and side/auxiliary information for a UE-side model are implementations-specific choices and cannot be pre-determined/standardized.* *Observation 3: The auxiliary/side information collected for the model development can be proprietary. Therefore, the data collected from/by a UE vendor should not be shared with other UE vendors, network vendors, operators (without service level agreement between operators and UE vendors), or third parties.* *Proposal 8: 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 4: 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 9: Considering the implementation-specific nature of the model input/output and auxiliary/side information and considering the runtime constraints (as mentioned in observation 4), the UE-side model can only be trained by the UE vendor, at least in the Rel-19 and foreseeable near future.* |
|  |  |

#### **Background**

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

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

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

1a) OTT (3GPP transparent)

1b) OTT (non-3GPP transparent)

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

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

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

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

## 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 |
| Panasonic | Support. |
|  |  |
|  |  |
|  |  |
|  |  |

# Model transfer/delivery

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 11: Support Case y as the baseline for model transfer/delivery for R19. Case z4 can be further studied when Case y is not able to serve the purpose.* |
| Ericsson[2] | *Proposal 10 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 11 Only if the collaboration burden of case y with NW-sided training is deemed infeasible, prioritize case z4 with specified model structure and coefficient precision.**Proposal 12 Conclude, from RAN1 perspective, that the model transfer/delivery Case z2 is deprioritized also for UE-part of two-sided model in Rel-19 due to the following reasons:* *•Risk of proprietary design disclosure* *•Burden of offline cross-vendor collaboration* |
| Huawei[3] | *Observation 5: 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 6: For model transfer/delivery where the model is trained at UE side or neutral site, the necessity of introducing Case z1 as opposed to the implementation manner of Case y is not clear:** Case z1 incurs the burden of offline cross-vendor collaboration, compared to Case y.** Case z1 may come with 3GPP NW side burden on model maintenance/storage compared to Case y.** Case z1 does not bring benefits compared to Case y.**Proposal 11: For model transfer/delivery where the model is trained at UE side or neutral site, assume Case y as the baseline.* |
| Intel[4] | *Observation 6:**• Collaboration level y offers a basic method for collaboration between UE and the network with limited specification impact but reduced efficiency due to reliance on offline coordination and model delivery as against over-the-air model transfer in use-cases involving site-/scenario-/configuration-specific models.**Observation 7:**• Offline model compilation and offline model testing, while desirable in general from perspective of UE implementation and model robustness, may not always be essential or justified considering the adverse impact to incurred latency for model updates and/or switching, e.g., for cases wherein model may be updated with respect to limited number of parameters while maintaining the model structure.**Proposal 6:**• From RAN1 perspective, model transfer/delivery Case z1 is deprioritized in Rel-19 due to the following reasons:**o Not much benefit compared to Case y.**o Large burden of offline cross-vendor collaboration.**o Additional burden on model storage within in 3GPP network.**o Limited applicability to only scenarios involving two-sided models with model transfer/delivery from UE to NW side.**Proposal 7:**• In Rel-19, consider support of model transfer/delivery Case y and model transfer/delivery Case z4 for model/parameter transfer/delivery.**o For model transfer/delivery Case z4, consider specifying a group/family of model structures/backbones to alleviate the burden of offline inter-vendor collaboration to align on model structure between NW and UE.* |
| Spreadtrum[5] | *Proposal 2: From RAN1 perspective, the model transfer/delivery Case z1 is deprioritized in Rel-19.**Proposal 3: From RAN1 perspective, the model transfer/delivery Case z2 is deprioritized in Rel-19 for two-sided model.**Proposal 4: Suggest to defer the discussion on Case z4 until good progress on AI9.1.3.2 multi-vendor issue achieved.* |
| IDC[6] | *Observation 11: In cases where model generalization, model finetuning or model storage/switching is not feasible, model delivery/transfer can be beneficial.**Proposal 15: Model transfer for UE-side models with functionality-based LCM is not supported and 3GPP specification transparent model delivery is only considered.* |
| Samsung[7] | *Proposal#8: Deprioritize study on Case z1 of 3GPP non-transparent model transfer, as it requires offline cross-vendor collaboration.* *Observation#3: 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#9: Study the feasibility and potential benefits of model (parameter) transfer for specified model structure from gNB to UE, i.e., Case z4.* |
| vivo[8] | *Proposal 11: Conclude that model transfer in open format of a known model structure at UE (i.e., Case z4) is feasible from device implementation perspective.**Observation 11: The burden of model storage would be relieved if the model structure is specified in 3GPP.**Observation 12: 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 13: From initial results for field test, cell/site specific model can provide up to 17.6% SGCS gain.**Observation 14: Field test shows model developed for one cell does not perform well for the other cell.* *Observation 15: Field test shows that simple and small models work well for different cases, at least for cell/site specific model.* *Proposal 12: Defining reference model (structures) is also beneficial from RAN4 testing perspective.* *Proposal 13: Support model transfer with known model structure at UE (Case z4).**Proposal 14: 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 15: Based on the specified model structure of reference model, model transfer with known model structure (Case z4) can be used to update model parameters.**Proposal 16: 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.* |
| Apple[9] | *Observation 1: Standardized model transfer solution for case z1 for UE side model has the following pros/cons:**Pros: lower latency for model delivery.* *Cons:* *• The burden of offline cross-vendor collaboration to send the trained model from the UE-side to the network.* *• The burden of model maintenance/storage on network side.* *Proposal 4: The necessity of standardized model transfer solution for case z2 and z4 can be further discussed after CSI compression use case is concluded.*  |
| CATT[10] | *Proposal 10: Further study model transfer z4 in Rel-19.**- Further study the value of z4 in terms of reduced effort from inter-vendor offline coordination, compared to case y with NW-side training;**- As a starting point, a small set of simple model structures can be considered as reference model structures;**- Further study the feasibility of parameter update with known model structure in UE device.* |
| CMCC[11] | *Proposal 8: 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 9: It is suggested to further study model transfer/delivery Case z4, from the following aspects:**• How to standardize reference model structure**• How to exchange model parameters**• The associated procedure* |
| NVIDIA[13] | *Proposal 4: Continue to study Cases y, z1 and z4 for transferring/delivering AI/ML model(s).* |
| LGE[14] | *Proposal#8. 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.* |
| Fujitsu[15] | *Observation-8: 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-11: Further study case z1 and case z4 for two-sided model only, including clarify the necessity to standardize model transfer/delivery from the following aspects at least:**• The performance benefits over using case y.**• The necessity of using case z1 and case z4.**• The feasibility of case z1 and case z4.* |
| Xiaomi[16] | *Observation 1: For the model trained by UE side or neutral site, the need to consider standardised solutions for transferring/delivering AI/ML model(s) is weak.**Observation 2: It is beneficial to support that AI models are trained by the network and then delivered/transferred to UE.**Observation 3: For Case y with NW side training** Large offline-coordination effort is required* * Large time-scale for model update* * Potential specification effort on the assistance signalling/procedure for the model transfer/delivery is necessary**Observation 4: For case z4, the following two options are possible for the model structure alignment between NW and UE** Option 1: Via offline coordination** Option 2: Via specified reference model* *Observation 5:* * For Case z4 with offline coordinated model structure, offline co-ordination effort is required* * For Case z4 with specified reference model, additional specification effort is required. But on the other hand, it could further facilitate the test for RAN4**Proposal 1: Consider standardised solutions for model transfer/delivery at least for the case that AI models are trained on network side.**Proposal 2: When the AI models are developed by the network side, prioritize investigating model transfer/delivery solution case z4 with specified model structure* |
| NEC[17] | *Observation 5: 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.**Proposal 13: Selection of model transfer methodology(ies) between z1 and z4 should be based on the outcome of CSI compression model transfer discussion.* |
| ZTE[19] | *Proposal 5: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer study for two-sided model rather than UE-side model.* *Observation#7: the overall prioritization up to RAN1#116bis is of the following.*

|  |  |  |
| --- | --- | --- |
|  | UE-sided model | Two-sided model |
| Case y |  |  |
| Case z1 |  |  |
| Case z2 | Deprioritized (RAN1#116bis) |  |
| Case z3 | Deprioritized (RAN1#116bis) | Deprioritized (RAN1#116bis) |
| Case z4 |  |  |
| Case z5 | Deprioritized (RAN1#116) | Deprioritized (RAN1#116) |

*Proposal 6: In Rel-19 AI/ML framework study, RAN1 prioritizes the model transfer z4 for two-sided model.* |
| Panasonic[20] | *Proposal 4: Model delivery/transfer cases comparison should be concluded as the following in RAN1.**- If training is UE-side, model can be sufficient to use proprietary format. Storage location can be up to other WG discussion.**- If training is NW-side, model need to be open format and stored within 3gpp network.* |
| OPPO[23] | *Proposal 10: 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.* |
| Nokia[24] | *Proposal 2: From RAN1 perspective, the model transfer/delivery Case z1 is deprioritized for Rel-19.**Proposal 3: For 2-sided CSI compression, particularly training type 1 (joint model training and model transfer/delivery to the UE), model transfer to be realized as user plane data transfer, controlled by the gNB/RAN using the control plane signaling.**Proposal 4: The RAN1 related configurations to be considered for the transfer should be flexible enough to allow full or partial model updates, i.e. the CP config might indicate this, and the transfer needs to include UE-vendor specific meta information.**Proposal 5: RAN1 to continue the study of the potential model transfer/delivery case z4 solutions for 2-sided CSI feedback enhancement (particularly for model training collaboration type 1) based on the agreed performance requirements.* |
| AT&T[25] | *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.* |
| DOCOMO[26] | *Proposal 4: Deprioritize case z1, unless explicit gain of case z1 compared to case y with UE side training is observed.* |
| Qualcomm[27] | *Proposal 10: For any UE-side model delivered/transferred to the target UEs for inference, 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. Any UE-side model delivery/transfer for inference that does not meet this requirement should be deprioritized.**Proposal 11: From RAN1 perspective deprioritize model transfer Case z4 for UE-side models, at least due to feasibility issues as well as lack of necessity to train UE-side models at the NW side.**Proposal 12: Add the following case to the existing model transfer/delivery cases:**• Case z0: model is trained at UE-side (i.e., at a UE-side server) and the model transfer happens from UE-side server (hosted at 3GPP network) to UE in a 3GPP non-transparent manner.* |
| IIT[29] | *Proposal 1: Further study Case Y and Case Z4 in Rel-19 for Model transfer/Delivery.* *Proposal 2: Deprioritize case Z1 for Rel-19 due to the following reasons:**• Risk of proprietary design disclosure**• Burden of offline cross-vendor collaboration.**Proposal 3: Deprioritize the model transfer/delivery Case z2 also for UE-part of two-sided model in Rel-19 due to the following reasons:**•  Risk of proprietary design disclosure**•  Burden of offline cross-vendor collaboration.* |
|  |  |

#### **Background**

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

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

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Case** | **Model delivery/transfer** | **Model storage location** | **Training location** |
| **y** | model delivery (if needed) over-the-top. | Outside 3GPP Network | UE-side / NW-side / neutral site |
| **z1** | model transfer in proprietary format. | 3GPP Network | UE-side / neutral site |
| **z2** | model transfer in proprietary format. | 3GPP Network | NW-side |
| **z3** | model transfer in open format. | 3GPP Network | UE-side / neutral site |
| **z4** | model transfer in open format of a *known model structure* at UE, i.e., an exact model structure as has been previously identified between NW and UE and for which the UE has explicitly indicated its support.  | 3GPP Network | NW-side |
| **z5** | model transfer in open format of *an unknown model structure* at UE, i.e., any other model structure not covered in z4, including any model structure that is only partially known. | 3GPP Network | NW-side |
| Note: The definition of various Cases is only for the purpose of facilitating discussion and does not imply applicability, feasibility, entity mapping, architecture, signalling nor any prioritization. |

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

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

## 1st round discussion

#### **Proposal 4.1.1**

* Case y: No spec impact on the air interface
* Case z1: During the last meeting, some new thing was coupled here.
* Case z2 for two-sided model: Please see Proposal 4.1.4.
* Case z4: Based on the tdocs, many companies suggest to clarify the procedures of Case z4. Thus, the following proposal is suggested for further discussion.

**Proposal 4.1.1**

**From RAN1 perspective, for model delivery/transfer Case z4, further study the following alternatives (including the necessity/benefits):**

* **Alt. A**
	+ **Step A-1: UE/UE-side reports to NW the supported known model structure(s)**
	+ **Step A-2: NW transfers to UE/UE-side the parameters and the associated model ID(s), which are corresponding to one or more of supported known model structure(s) reported in Step A-1**
* **Alt. B**
	+ **Step B-1, NW indicates to UE/UE-side the candidate known model structure(s)**
	+ **Step B-2, UE reports to NW which model structure(s) out of the candidate known model structure(s) indicated in Step B-1 is supported**
	+ **Step B-3, NW transfers to UE/UE-side the parameters and the associated model ID(s), which are corresponding to one or more of supported known model structure(s) reported in Step B-2**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | We agree there can be two candidates as described alt A and alt B. On the other hand, Alt B is required to support multiple of model structures. In order to limit the candidates of model structure, we rather think alt A is more reasonable. If there are multiple of the candidates of the models are required, these are listed in UE capability signaling in step A-1. |
| Huawei, HiSilicon | 1) Generally fine with two directions (NW proactive and UE proactive). But, compared with the procedure, isn’t it more imperative to start discussing how to converge on the model structure itself in 3GPP? In our understanding, we may have a long way to go before having consensus on the exact neuron by neuron level model structure, yet there are only two meetings before Sept.2) The intention of aligning model structure is to enable a plug-and-play manner at UE – as soon as UE receives the parameters from NW, it can almost immediately map the parameters to the run time image and run the model.If NW anyway needs to deliver the model parameters to the UE side server for a further retraining/updating, which implies that the model/model structure used by UE device is still a proprietary structure and probably different from the specified model structure, then why do we still waste spec effort to specify the model structure?**From RAN1 perspective, for model delivery/transfer Case z4, further study the following alternatives (including the necessity/benefits):*** **Alt. A**
	+ **Step A-1: UE~~/UE-side~~ reports to NW the supported known model structure(s)**
	+ **Step A-2: NW transfers to UE~~/UE-side~~ the parameters and the associated model ID(s), which are corresponding to one or more of supported known model structure(s) reported in Step A-1**
* **Alt. B**
	+ **Step B-1, NW indicates to UE~~/UE-side~~ the candidate known model structure(s)**
	+ **Step B-2, UE reports to NW which model structure(s) out of the candidate known model structure(s) indicated in Step B-1 is supported**
	+ **Step B-3, NW transfers to UE~~/UE-side~~ the parameters and the associated model ID(s), which are corresponding to one or more of supported known model structure(s) reported in Step B-2**
 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 4.1.2**

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

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

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

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

**Proposal 4.1.2**

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Panasonic | Support |
| Huawei, HiSilicon | OK |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 4.1.3**

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

**Proposal 4.1.3**

**For the open format for model delivery/transfer Case z4, further study the following Options (including the feasibility/specification efforts)**

* **Option 1: Reuse the existing open format(s) that has existed in the AI community (e.g., ONNX)**
	+ **FFS: which open format(s)**
* **Option 2: Define a new open format within 3GPP**

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Mod | We may remove/add some option (depending on the inputs/comments) |
| Panasonic | Support. Option 1. |
| Huawei, HiSilicon | This proposal is dependent on Proposal 4.1.2, so it can be deferred after 4.1.2 is concluded. If 4.1.2 is agreed, then we do not need to discuss the model representation format – using ASN.1 is sufficient to represent the quantized parameters to be transferred from NW to UE. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

#### **Proposal 4.1.4**

* In the last meeting, most companies expect one agreed to deprioritize Case z2 for both two-sided and one-side model
* In the submitted tdocs, most companies continue proposing to deprioritize Case z2 for two-sided model
* @QC: would you like to elaborate which option(s) (e.g., 3a? 5a?) is related to Case z2 and how it is related to Case z2?

**Proposal 4.1.4**

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

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

Companies can provide comments/inputs in the following table:

|  |  |
| --- | --- |
| Company | Comment |
| Moderator | @QC: Would you like to elaborate which option(s) (e.g., 3a-1? 3a-2? 3a-3? 3b?5a-1? 5a-2? 5a-3?) is related to Case z2 and how it is related to Case z2? Not sure whether the following Option X1 for Case z2 for two-sided model is considered from QC side:* Option X1: model transfer in proprietary format with a known model structure at UE
* Option X2: model transfer in proprietary format with an unknown model structure at UE

In moderator’s understanding, Option 1/2/4/5 of CSI compression seem NOT related to Option X1) and Option 1/2/3/4/5 of CSI compression seem NOT related to Option X1. Thus, the only possible link is between Option X1 and Case 3a/3b. Is that the correct understanding? @all companies: we will not discuss this proposal until QC makes some clarification so that the group can decide whether we need to wait for progress of CSI compression or not. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Others

#### **Companies’ view**

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| CATT[10] | *Observation 10: 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 12: For model-ID-based LCM (if supported), it is beneficial to assess/monitor the performance of a specific inactive model at UE-side.**Proposal 13: For functionality-based LCM, it is beneficial to assess/monitor the performance of a specific inactive functionality at UE-side.**Proposal 14: 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.* |
| NVIDIA[13] | *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.* |
| LGE[14] | *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[17] | *Observation 4: 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.**Observation 6: It is important to discuss how UE can indicate its internal restrictions to activate or run an AI/ML model/functionality to the network for optimal AI/ML operation.**Observation 7: Reporting of UE’s internal conditions such as memory size, battery level and other detailed hardware limitations to gNB for AI/ML operation may lead to UE’s proprietary information disclosure and may be hard for network to determine AI/ML applicability for a UE based on the provided information.**Proposal 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.**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 14: Specify UE indication to network about its inability to run a configured/activated AI/ML model/functionality due to UE’s internal condition along with a relevant cause value for the failure.* |
| AT&T[25] | *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 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.* |
| Continental Automotive[28] | *Proposal 6: 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 7: Study of online training is suggested.**Proposal 8: Study of UE ML capability related to training collaboration aspect is suggested.* |
|  |  |

**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 Monday’s online session

# Appendix A: Agreements

## RAN1#116

Agreement

* To facilitate the discussion, RAN1 studies the model identification type A with more details related to use cases.
* To facilitate the discussion, RAN1 studies the following options as starting point for model identification type B with more details related to all use cases
* MI-Option 1: Model identification with data collection related configuration(s) and/or indication(s)
* MI-Option 2: Model identification with dataset transfer
* MI-Option 3: Model identification in model transfer from NW to UE
* FFS: The boundary of the options
* Note: the names (MI-Opton1, MI-Option 2, MI-Option 3) are used only for discussion purpose
* Note: other options are not precluded

**Observation**

The other options are proposed for model identification type B by companies during the discussion:

* MI-Option 4. Model identification via standardization of reference models. (for CSI compression)
* MI-Option 5. Model identification via model monitoring

Agreement

* Regarding MI-Option 1 (Model identification with data collection related configuration(s) and/or indication(s)) of model identification type B, RAN1 further study the following aspects:
* Relationship between model ID and data collection related configuration(s) and/or indication(s)
* Information transmitted from NW to UE (if any)
* Information transmitted from UE to NW (if any)
* The associated procedure
* Usage/Applicable use case(s) of MI-Option 1

Note: whether MI-Option 1 is needed or not is a separate discussion

**Conclusion**

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

**Conclusion**

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

## RAN1#116bis

Conclusion

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

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

Conclusion

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

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

Conclusion

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

Agreement

From RAN1 perspective, for UE-sided model(s) developed (e.g., trained, updated) at UE side, following procedure is an example (noted as **AI-Example1**) of MI-Option1 for further study (including the feasibility/necessity)

* A: For data collection, NW signals the data collection related configuration(s) and it/their associated ID(s)
	+ Associated IDs for each sub use case in relation with NW-sided additional conditions
* B: UE(s) collects the data corresponding to the associated ID(s)
* C: AI/ML models are developed (e.g., trained, updated) at UE side based on the collected data corresponding to the associated ID(s).
* D: UE reports information of its AI/ML models corresponding to associated IDs to the NW. Model ID is determined/assigned for each AI/ML model
	+ relationship between model ID(s) and the associated ID(s)
	+ How model ID(s) is determined/assigned, e.g.,
		- Alt.1: NW assigns Model ID
		- Alt.2: UE assigns/reports Model ID
		- Alt.3: Associated ID(s) is assumed as model ID(s)
			* “Model ID is determined/assigned for each AI/ML model” in D is not needed
		- Alt.4: Model ID is determined by pre-defined rule(s) in the specification
	+ FFS: how to report
	+ Note: D is to facilitate AI/ML model inference
* Note: Step A/B/C and additional interaction of associated IDs between UE and NW can be considered as a different solution for resolving the consistency without model identification.

# 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 YerrapragadaSai Prasad Pirati | venkatasiva@5gtbiitm.in |
| CATT | Yongqiang Fei | feiyongqiang@catt.cn |
| Baicells | Xiang YUNXiaonan WANG | yunxiang@baicells.comwangxiaonan@baicells.com |
| Lenovo | Jianfeng WangVahid Pourahmadi | wangjf20@lenovo.comvpourahmadi@lenovo.com |
| ZTE | Xingguang WEI | 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 WangShijia Shao | Hualei.wang@unisoc.comShijia.shao@unisoc.com |
| AT&T | Isfar Tariq | Isfar.tariq@att.com |
| CMCC | Yuhua CaoYi ZhengDan Song | caoyuhua@chinamobile.comzhengyi@chinamobile.comsongdan@chinamobile.com |
| NEC | Peng GuanPravjyot Singh Deogun | guan\_peng@nec.cnpravjyot.deogun@EMEA.NEC.COM |
|  |  |  |
| CEWiT | Ebin ChackoShiv Shankar | echacko@cewit.org.inshivshankar@cewit.org.in |
| Intel | Debdeep Chatterjee | debdeep.chatterjee@intel.com |
| NVIDIA | Xingqin Lin | xingqinl@nvidia.com  |
| CICTCI | Rui Zhao | zhaorui@cictci.com |
| Continental Automotive | Hojin Kim | hojin.kim@continental-corporation.com  |
| LG Electronics | Jiwon Kang | jw.kang@lge.com |
| LG Electronics | Haewook Park | haewook.park@lge.com |
| Futurewei | Chunhui (Allan) Zhu | chunhui.zhu@futurewei.com |
| MTK | PedramYu-Jenharrison | pedram.kheirkhah@mediatek.comyu-jen.ku@mediatek.comharrison.chuang@mediatek.com |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| Mavenir | Fan YangYuanlong Yang | fan.yang@mavenir.comyuanlong.yang@mavenir.com |
| Ruijie | Ke Zhong | zhongke@ruijie.com.cn |

# Appendix B: Reference/tdocs

1. R1-2403868 Discussion on other aspects of AI/ML model and data on AI/ML for NR air-interface FUTUREWEI
2. R1-2403915 Discussion on other aspects of AI/ML Ericsson
3. R1-2403933 Discussion on other aspects of the additional study for AI/ML Huawei, HiSilicon
4. R1-2403977 Other study aspects of AI/ML for air interface Intel Corporation
5. R1-2404018 Discussion on other aspects of AI/ML model and data Spreadtrum Communications
6. R1-2404055 Discussion on other aspects of AI/ML model and data InterDigital, Inc.
7. R1-2404105 Discussion for further study on other aspects of AI/ML model and data Samsung
8. R1-2404169 Other aspects of AI/ML model and data vivo
9. R1-2404276 Discussion on other aspects of AI/ML model and data Apple
10. R1-2404388 Study on AI/ML for other aspects of model and data CATT, CICTCI
11. R1-2404448 Discussion on other aspects of AI/ML model and data CMCC
12. R1-2404529 On aspects of AI/ML model and data framework Lenovo
13. R1-2404540 Additional study on other aspects of AI model and data NVIDIA
14. R1-2404549 Discussion on other aspects of AI/ML model and data LG Electronics
15. R1-2404586 Discussion on other aspects of AI/ML model and data Fujitsu
16. R1-2404605 Further study on AI/ML model and data Xiaomi
17. R1-2404656 Discussion on other aspects of AI/ML model and data NEC
18. R1-2404686 AI/ML Model and Data Google
19. R1-2404704 Discussion on study for other aspects of AI/ML model and data ZTE
20. R1-2404756 Discussion on other aspects for AI/ML for air interface Panasonic
21. R1-2404764 View on AI/ML model and data MediaTek Korea Inc.
22. R1-2404769 Discussion on other aspects of AI/ML model and data ETRI
23. R1-2404881 Additional study on other aspects of AI/ML model and data OPPO
24. R1-2404908 Other Aspects of AI/ML Model and Data Nokia
25. R1-2405017 Other Aspects of AI/ML framework AT&T
26. R1-2405034 Discussion on other aspects of AI/ML model and data NTT DOCOMO, INC.
27. R1-2405147 Other aspects of AI/ML model and data Qualcomm Incorporated
28. R1-2405212 Discussion on other aspects of AI/ML model and data Continental Automotive
29. R1-2405304 Discussion on other aspects of AI/ML model and data IIT Kanpur, Indian Institute of Tech (M).