**3GPP TSG-RAN WG2 Meeting #121-bis electronic R2-230xxxx**

**April 17-26, 2023**

**Agenda Item:** **7.16.2.1**

**Source: OPPO (email rapporteur)**

**Title: Report of [AT121bis-e][014][AIML18] Model ID (incl meta data) progress (OPPO)**

**Document for: Discussion and Decision**

# Introduction

This is the offline report of [AT121bis-e][014]:

* [AT121bis-e][014][AIML] Model ID (incl meta data) progress (OPPO)

Scope: Take into account relevant input to this meeting. Determine the use cases and usefulness of Model ID, potential additional meta data.   
Collect Comments, Identify easy agreements (if any), potential agreements, and Open Issues (which seem important to address). Pave the way for online Come-Back

Intended outcome: Report

Deadline: Schedule 1

**Discussions with Deadline Schedule 1:**

**A first round with Deadline W1 Friday April 21th 1000 UTC to settle scope what is agreeable etc (at latest, Rapp may also set an earlier deadline)**

**General timeline guidance for this offline discussion:**

* **Comment deadline: Friday W1, 1000 UTC (for collecting views)**
* **Rapporteur summary deadline: Monday W2, 0600 UTC (proposed outcome)**
* **Offline document upload deadline: 1h before Monday W2 online session (discussion report)**

In this offline discussion, it’s suggested to focus on model ID and meta data topics. For model ID, it is suggested to discuss the use cases first and try to agree a list of cases for which a model ID shall/should be used. Then we will try to clarify the mechanism on how “globally” unique model ID works and whether to have more ID types. As for meta data, it is also suggested to first discuss the typical use cases and try to agree a list of cases for which a meta data can be useful. Then we will try to have an initial discussion on the contents of meta data to facilitate our future study.

**Note1: In the offline discussion, there is no intention to discuss the boundary/necessity between model identification and functionality identification (or between model management and functionality management), whether to have one or both functions depends on RAN1 inputs. Companies should focus on the assumption that model -ID-based LCM is considered and give your technical comments.**

**Note2: In this offline discussion, considering a specific LCM purpose does not imply the necessity to introduce the corresponding function, whether to introduce the corresponding function should be discussed separately, which is out of the scope of this offline discussion. In this offline, companies are suggested to focus on how/why model ID and meta data should be involved in LCM procedures.**

Companies providing input to this offline discussion are requested to leave contact information below.

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| Huawei, HiSilicon | Jun Chen | jun.chen@huawei.com |
| vivo | Boubacar Kimba | kimba@vivo.com |
| Apple | Peng Cheng | pcheng24@apple.com |
| Ericsson | Felipe Arraño Scharager | felipe.arrano.scharager@ericsson.com |
| T-Mobile USA | John Humbert | John.Humbert2@T-Mobile.com |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Discussion

## 2.1 Discussion on model ID

### 2.1.1 Use case discussion for model ID

Model ID topic has been widely discussed in the past several meetings across RAN1 and RAN2, although there is no explicit LS exchange on the usage of model ID between working groups, the identified use cases for model ID are almost aligned with each other so far except fallback use case, so before discussing new use case for model ID, it’s better to first have common understanding on the agreements made so far for model ID across RAN1 and RAN2, the background details are given below:

In RAN2#119bis\_e meeting, RAN2 assumes that a model is identified by a model ID, but the use case is FFS [3]:

* *R2 assumes that a model is identified by a model ID. Its usage is FFS.*

In RAN2#120 meeting in Toulouse, RAN2 further agreed that model ID can be used to identify a model (or models) during model selection/activation/deactivation/switching [2]:

* *R2 assumes that model ID can be used to identify a model (or models) during model selection/activation/deactivation/switching (can later align with R1 if needed).*

During the same meeting period in Toulouse, RAN1 also had a similar agreement on the usage of model ID in RAN1#111 meeting [5]:

*Agreement*

*For UE-part/UE-side models, study the following mechanisms for LCM procedures:*

* *For functionality-based LCM procedure: indication of activation/deactivation/switching/fallback based on individual AI/ML functionality*
  + *Note: UE may have one AI/ML model for the functionality, or UE may have multiple AI/ML models for the functionality.*
  + *FFS: Whether or how to indicate functionality*
* *For model-ID-based LCM procedure, indication of model selection/activation/deactivation/switching/fallback based on individual model IDs*

Based on above agreement, it seems that RAN1 initially had intention to add fallback as one of the use cases for model ID. But in RAN1#112 meeting, fallback is not explicitly added as one of the use cases for model-ID-based LCM [7]:

*Agreement*

*For UE-side models and UE-part of two-sided models:*

* *For AI/ML functionality identification*
  + *Reuse legacy 3GPP framework of Features as a starting point for discussion.*
  + *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.*
* *In functionality-based LCM*
  + *Network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signaling (e.g., RRC, MAC-CE, DCI).*
  + *Models may not be identified at the Network, and UE may perform model-level LCM.*
    - *Study whether and how much awareness/interaction NW should have about model-level LCM*
* *In model-ID-based LCM, models are identified at the Network, and Network/UE may activate/deactivate/select/switch individual AI/ML models via model ID.*

*FFS: Relationship between functionality identification and model identification*

*FFS: Performance monitoring and RAN4 impact*

*FFS: detailed understanding on model*

Although RAN1 agreements were not consistent across meetings, at least, RAN1 and RAN2 can reach common understanding that model ID is used at least for model selection/activation/deactivation/switching, so we propose the Observation1:

**Observation1: For model-ID-based LCM, model ID can at least be used to identify a model (or models) during model selection/activation/deactivation/switching.**

To be safe, companies are invited to share views on Q1. Please also take the Notes given in the introduction section into account when you have comments.

**Q1: Do companies agree the Observation1 above? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | We think ob1 is the previous RAN2 working assumption and we have not seen concerns from companies, so it should be agreeable. |
| Xiaomi | Yes |  |
| NEC | Yes |  |
| vivo | Yes |  |
| ZTE | Yes | This is a previous agreement. |
| Apple | Yes | Observation 1 use wording "can", then "at least" is redundant. Suggest to remove "at least". |
| Ericsson | Yes | Plus, we would like to highlight the importance of referring to model**s** as the model ID design could end up pointing to a set of models intended for e.g. a functionality. |
| T-Mobile USA | Partially | Yes for model selection/activation/deactivation/switching. however, we strongly disagree with the concept that a model ID can refer to more than one model. Having a model ID refer to multiple models severely impacts MNO’s ability to test, verify , monitor AI/ML impact on the network. |

**Placeholder for summary**:

As for the fallback use case, from offline rapporteur’s point of view, it may be hard for RAN2 to make the judgement that this use case is missed by RAN1 or is clearly ruled out by RAN1 on purpose, so the safer way is to leave this use case FFS for now. Anyway, three options are given for comments:

**Option1**: For model-ID-based LCM, model ID is not used for fallback LCM purpose.

**Option2**: For model-ID-based LCM, model ID can be used for fallback LCM purpose.

**Option3**: For model-ID-based LCM, it’s FFS whether model ID is used for fallback LCM purpose.

Companies are invited to share views on Q2. Please also take the Notes given in the introduction section into account when you have comments.

**Q2: Which Option is preferred from your side? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Option1/2/3** | **Comments** |
| Huawei, HiSilicon | Option 3 | In our understanding, fallback means a change from AI-based state to non-AI-based state, e.g. initially the UE is using AI-based CSI prediction, and then due to some reasons, the network does not want to enable this AI based solution, so the UE can be configured to go to legacy CSI handling (without any AI-based CSI prediction).  For now, we think it is hard for RAN2 to make a decision, and it rely on the meaning/requirement of fallback (may need more RAN1 inputs), so we prefer option 3. |
| Xiaomi | 3 | There is no common understanding about the need of fallback procedure. It’s too early to conclude whether model ID is needed for fallback. |
| NEC | 3 | We prefer RAN1 to confirm the need to take this fallback use case for AIML operation. |
| vivo | Option 1 | The network may deactivate the AI model with model ID(s) and indicate the UE to fallback to legacy behavior without extra model ID. |
| ZTE | Option 3 | It is still ambiguous for us about what the is ‘fallback’. From RAN2 perspective, we just know that the fallback means that UE is forced to be switched from AI based to non-AI based for one specific feature, it is feature granularity not model granularity.  And for now, the most safe way is option 3. |
| Apple | 3 (leave it to RAN1) | Same view as Huawei. If fallback is chosen by UE/NW, legacy signaling is used and model ID should not be used. Whether some specifical model ID is used to indicate fallback mode is stage 3 issue, which should not be discussed in SI phase.  So, although we have sympathy with option 1, the conclusion should be made in RAN1 because how LCM work is led by RAN1. |
| Ericsson | Option 3 | Agree with HW. |
| T-Mobile USA | Option 1 | Agree with Vivo’s comment, if all model ID’s are disabled for a function then legacy functionality should be used. |

**Placeholder for summary**:

Regarding to model transfer/delivery LCM purpose, several companies think model ID can be used for this LCM purpose [9][10][20][21], the following RAN2 agreement is also raised as one of the reasons [2]:

* *R2 assumes that model ID can be used to identify which AI/ML model is being used in LCM including model delivery.*

More addition, the proponents also think model ID should be linked to the corresponding model algorithm for subsequent model management, e.g. model activation/deactivation. The key difference among proponents is the signaling to link a model ID and the corresponding model algorithm, for instance, CP or UP signaling. From offline rapporteur perspective, our main target in this offline is to address the use cases for model ID, the signaling can be discussed in the future meeting based on contributions, so only one thing that should be confirmed in this offline is whether mode ID can be used for model transfer/delivery LCM purpose.

Companies are invited to share views on Q3. Please also take the Notes given in the introduction section into account when you have comments.

**Q3: Do companies agree that model transfer/delivery LCM purpose is one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | During model transfer/delivery procedure, we think it should be possible for the training entity (or the entity which stores the models) to send the model info together with model ID to the UE. It is also possible that the training entity only sends model info to the UE (without any model ID), and then it may take more efforts for UE and network to synchronize on the model ID (maybe other info), but this can be further discussed. |
| Xiaomi | Yes with comments | We understand the model ID can be indicated together with the actual model during model delivery. However, we don’t agree that only model ID is indicated without the actual model during model delivery. Because the UE may not be able to store all the models mapped by model ID. |
| NEC | Yes | We think there may be a case where the network will not send model ID together with the model itself to the UE, e.g. if UE asks the model with a ID and the network may transfer the model to UE accordingly. |
| vivo | Yes | For the model transfer from NW to UE, each model shall associate with one model ID for further LCM purposes, e.g., model update. |
| ZTE | Yes | Model Id, which is model unique. And in a common sense, the model transfer should support dedicate transferring of one model from a side to another. So the model Id is a good means to make two endpoint of the model transfer have a unified understanding on the AI/ML Model for transferring. |
| Apple | Yes | We think there are below 4 possible signaling to identify a model during model transfer:   1. Model ID + meta info + actual model 2. Model ID + actual model 3. Model ID + meta info 4. Model ID   The difference between 1) and 2) is whether it is necessary to also provide optional model meta info to align understanding between UE and NW.  We think case 3) and 4) is also useful because the UE may locally store some model or the UE prefer to download from its own server. So, based on model ID + meta info. the UE can determine whether/which mode needs delivery from the NW. |
| Ericsson | No, see comment | RAN2 have not yet agreed to supporting model transfer/delivery. Hence, we can/should only keep it as an assumption for now. |
| T-Mobile USA | Yes | Agree with Vivo’s comments |

**Placeholder for summary**:

Data collection LCM purpose is a generic topic which may involve several other LCM purposes based on RAN1 agreements made in RAN1#110b meeting [4]:

*Conclusion*

*Data collection may be performed for different purposes in LCM, e.g., model training, model inference, model monitoring, model selection, model update, etc. each may be done with different requirements and potential specification impact.*

*FFS: Model selection refers to the selection of an AI/ML model among models for the same functionality. (Exact terminology to be discussed/defined)*

Rapporteur suggests to focus on data collection for offline training in this section as the other LCM purposes involving data collection will be discussed later in the corresponding LCM purposes, i.e. data collection for model inference and model monitoring.

Although model input is defined per AI/ML model, there may be no need to use model ID during data collection procedure for offline training. Because some existing data collection frameworks, e.g. SON and MDT procedure, may already collect some types of data which can also be used for offline training even if the original data collection target is aimed for other optimization use cases raised by SA5. Even if new data collection framework is introduced for offline training procedure, it’s still possible that the collected data via new data collection framework can be shared by multiple models and/or other optimization purposes. Collecting data for model training per model ID is inefficient and not necessary. Based on above, to leave some flexibility for the usage of collected data, model ID should not be involved for data collection procedure for offline training.

Companies are invited to share views on Q4. Please also take the Notes given in the introduction section into account when you have comments.

**Q4: Do companies agree that data collection for offline training is not considered as one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Suggest to put FFS | We observe that RAN1 is also discusing “scenario ID”, “applicability ID” and others, which may lead to new requirements for RAN2. So it will be safe to keep it open for now and we may check it later once RAN1 has more progress. |
| Xiaomi | Comments | Similar view as HW, we may be careful about ruling out things at this stage, since RAN1 discussion is pretty diverse. |
| NEC | Postpone | May wait for outcome of the discussion at Ran1 |
| vivo | Yes | RAN2 may assume no associated model ID is needed during the data collection for offline model training. |
| ZTE | FFS | Similar view as HW. It is in the study item, no need to preclude the application of the model Id since we understand the model Id is more practical discussion point in WI stage. |
| Apple | FFS | Same view as Huawei and Xiaomi. We believe both RAN1 and RAN2 don't have aligned understanding on requirement of data collection for AI/ML at this stage. Not sure how RAN2 can make a conclusion for a feature without clear understanding.  We think one use case can be training collaboration type 3 with separate training at NW side and UE side, where both model and dataset may be delivered from one entity to the peer entity. For example, for UE-first training, UE will collect data, perform initial training, and generate the training dataset for NW training. The dataset can be transmitted together with the model ID label, so NW side can perform separate training using the received dataset and knows how to pair the UW side model with UE side model. |
| Ericsson | FFS | If we decide to go along the lines of using these IDs, then the data collection configuration could be linked to a particular family of IDs, for which there is an eventual dependency. Hence, better to have it as FFS. |
|  |  |  |

**Placeholder for summary**:

Data collection for offline training may or may not have spec impact which is addressed in Q4, but how to use the collected data for offline model training is assumed to be implementation domain issue, so apart from data collection for offline training, it seems there is no need to further consider model ID to be used in other aspects, e.g. model generation procedure including model validation and testing, of offline training.

Companies are invited to share views on Q5. Please also take the Notes given in the introduction section into account when you have comments.

**Q5: Is there any other aspect of offline training that should be considered as one of the use cases for model ID while data collection for offline training is addressed in Q4? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Apple | Yes | We think one missed use case is model pairing in two-sided model collaboration without model transfer (i.e. in training collaboration type 2 and type 3):   * For training collaboration type 2 with joint training of the two-sided model at network side and UE side jointly, the NW and UE need to align the understanding of the pair of used model ID, especially if NW and UE use different AI/ML vendors.   For training collaboration type 3 with separate training at NW side and UE side, model is needed to pair the AI/ML models between the UE and the network. For example, the UE can notify NW its model ID used for CSI compression, then NW can perform separate training using the received dataset and knows how to pair the NW side model with UE side model. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Placeholder for summary**:

In RAN1 discussion, model identification is used in the past two meetings instead of model registration terminology. A parallel terminology named functionality identification is also introduced for further discussion.

For model identification procedure, model ID is used to identify model, while for functionality identification procedure, model ID may or may not be used as RAN1 is still discussing the relationship between functionality identification and model identification. The related RAN1 agreements are given below for reference [7]:

*Agreement*

*For UE-side models and UE-part of two-sided models:*

* *For AI/ML functionality identification*
  + *Reuse legacy 3GPP framework of Features as a starting point for discussion.*
  + *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.*
* *In functionality-based LCM*
  + *Network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signaling (e.g., RRC, MAC-CE, DCI).*
  + *Models may not be identified at the Network, and UE may perform model-level LCM.*
    - *Study whether and how much awareness/interaction NW should have about model-level LCM*
* *In model-ID-based LCM, models are identified at the Network, and Network/UE may activate/deactivate/select/switch individual AI/ML models via model ID.*

*FFS: Relationship between functionality identification and model identification*

*FFS: Performance monitoring and RAN4 impact*

*FFS: detailed understanding on model*

Based on above sentence highlighted yellow, it seems that this RAN1 agreement implies that model ID is involved/used in model identification procedure. Several proponents also think model ID can be considered for model identification [9][12][14][15][20].

Companies are invited to share views on Q6. Please also take the Notes given in the introduction section into account when you have comments.

**Q6: Do companies agree that model identification is one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | We think RAN2 previously used model registration and later they used model identification instead (and functionality identification). Even if the wordings are different, the principle seems the same, i.e. after the process of identification, both UE and NW will have the common understanding on an AIML model.  As discussed in our paper [20], we think that for UE-sided/UE part models, model ID may be exchanged during model registration/identification, and we have some proposals (e.g. online/offline manner). So we think yes for Q6.  In addition, for network-sided models, we think it is unclear whether model identification is needed or not, and RAN2 may further discuss it. |
| Xiaomi | Yes | We understand the model identification is used to reach common understanding on the candidate model between UE and NW. UE and NW can exchange the global unique model ID of the candidate model, rather than the actual model. Model ID can reduce the exchange signaling between UE and NW during model identification. |
| NEC | Yes |  |
| vivo | Yes |  |
| ZTE | Yes |  |
| Apple | Yes with comments | We agree global unique model ID can be used in model identification, to ensure both UE and NW will have the common understanding on an AIML model.  However, we have different view on some companies' comments:  1) We disagree to use terminology "model registration", which is not agreed by RAN1. RAN2 should only use terminology of "model identification", to align with RAN1.  2) For "model identification", we think RAN1 is discussing whether it needs to be specified or it is up to offline implementation. So, we disagree to discuss procedure / signaling or model identification in RAN2 at this stage. RAN2 need to wait RAN1 conclusion. |
| Ericsson | Yes | We see these as being equivalent. |
| T-Mobile USA | Yes |  |

**Placeholder for summary**:

Another thing that should be noted is that it’s still unclear whether RAN1 would like to introduce either functionality identification or model identification or even have a combination solution. From RAN2 perspective, the safer way is to put functionality identification into FFS for the usage of model ID and wait for more RAN1 guidance.

Companies are invited to share views on Q7. Please also take the Notes given in the introduction section into account when you have comments.

**Q7: Do companies agree that more RAN1 inputs are still needed to judge whether to consider functionality identification as one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | Firstly, we think functionality identification is still under discussions in RAN1, so more RAN1 inputs will be helpful (but best efforts).  Secondly, our views on model/functionality identification is that, the two ways can be seprately discussed in order to make either work. And then we may consider the combination, but it may take more time.  As discussed in our paper [20], we think sub use case level can be considered and legacy 3GPP framework of features could be considered for functionality identification. In other words, a functionality can be considered as a feature, and then the relevant LCM procedures can be largely simplified compared with model-based LCM procedures, e.g. model/model ID related procedures may not be needed any more. |
| Xiaomi | Yes |  |
| NEC | FFS | We may need more input from RAN1 on the functionality identification before answering the question. |
| vivo | Yes |  |
| ZTE | Yes | RAN1 has put the FFS that the relationship between the model Identification and functionality identification. The conclusion of such FFS is a key factor to judge whether the model Id is coupled with the functionality identification or not. |
| Apple | Yes | RAN1 can't achieve consensus on below issue2:  1) Granularity of functionality based LCM  2) Relationship between model ID based LCM and functionality based LCM  Thus, RAN2 conclusion (if any) will only confuse RAN1. We suggest RAN2 to focus on model ID based LCM, and revisit functionality based LCM only when RAN1 conclusion is clear. |
| Ericsson | No | As we see it, RAN2 can already now separate these topics as distinct discussion. So perhaps RAN2 can already now agree that model-ID-based and functionality-based management are possible approaches.   Note that we are not saying that functionality IDs are needed. On the contrary, we think they are not and this could eventually be “business as usual” for RAN2 on how UEs indicate support for a functionality/feature. |
| T-Mobile USA | No | Functionality ID is an import part of AI/ML, current UE capabilities isn’t scalable to handle 10,000’s of model id’s. |

**Placeholder for summary**:

As for model inference, the terminology definition is given below for information [8]:

|  |  |
| --- | --- |
| AI/ML model Inference | A process of using a trained AI/ML model to produce a set of outputs based on a set of inputs |

Considering several other procedures are tightly associated with inference procedure defined above, all the following procedures should be analyzed:

Procedure1: data collection for model inference;

Procedure2: collected data pre-processing;

Procedure3: using a trained AI/ML model to produce a set of outputs based on a set of inputs;

Procedure4: Post-processing for model outputs.

It’s obvious that the above model inference definition given by RAN1 only covers Procedure3, it seems that model ID is not used for model inference procedure, i.e. Procedure3, but for the rest Procedures, it’s more suitable for RAN1 to discuss whether model ID is used for these Procedures, i.e. Procedure1/2/4. From offline rapporteur perspective, we can focus on Procedure3 in this offline discussion, the rest Procedures can be further clarified by RAN1.

Companies are invited to share views on Q8. Please also take the Notes given in the introduction section into account when you have comments.

**Q8: Do companies agree that model inference, i.e. A process of using a trained AI/ML model to produce a set of outputs based on a set of inputs, is not considered as one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | FFS | Too early to decide. More RAN1 inputs will be helpful. |
| Xiaomi | Yes | We understand the inference should be left to implementation. 3GPP would not specify how to make inference. But fine to keep it open in case RAN1 has related agreement. |
| NEC | FFS | We may need more input from RAN1 before answering the question.  Model ID can be useful during model inference, for example, where two sided model applied or more than one models are activated. |
| vivo | FFS | The request of input and report of output may associate with model ID. |
| ZTE | Comment | It is a really confusing question, is there any difference between the model inference and model activation?? To our understanding, to activate an AI model is equal to start the corresponding model inference. |
| Apple | Leave it to RAN1 | We also think this question is confusing because RAN2 don't touch model inference before. So, different companies have different understanding on what is model inference.  Because RAN1 is leading study model inference, we think the conclusion should be made in RAN1. RAN2 can just wait RAN1. |
| Ericsson | Yes | Agree with the intention, but uncertain whether we need to agree to this. Perhaps some further clarification is needed? In principle, when the inference process is being carried, one can assume that the model has been signaled/configured/activated/etc… |
| T-Mobile USA | No | Question is confusing. Model ID must refer to a single inference, if model outputs are allowed to change based on implementation, then there is no way to test or monitor the performance of the model in a real network. Any inference change needs to trigger a new model ID. |

**Placeholder for summary**:

When it comes to mode monitoring, several proponents also think model ID can be used for model monitoring [9][10][20][21], the following sub-use cases may be involved:

Case1: UE side model with UE side model monitoring without network side involvement;

Case2: Network side model with network side model monitoring without UE side involvement;

Case3: Two side model with network side model monitoring;

Case4: UE side model with network side model monitoring;

Case5: UE side model with UE side model monitoring with network side involvement.

For Case1 and Case2, it’s up to UE/NW implementation to handle the model monitoring procedure, so there is no need to consider model ID for Case1 and Case2.

But for the remaining Cases, i.e. Case3, Case4 and Case5, model ID may be used. For Case3 and Case4, network may need to collect some UE side metrics which can be used as part of the network side model monitoring inputs. Usually the UE side metrics should be collected per model, so model ID may be included into the model monitoring configuration and reports. For Case5, network side assistant info, e.g. specific reference signaling for, may be needed for UE side model monitoring and the network side assistant info may be configured to UE per model ID granularity. It should be noted that RAN1 is still discussing the details for model monitoring, to avoid any misalignment with RAN1, in this meeting, rapporteur suggests to put this use case into FFS part for the usage of model ID.

Companies are invited to share views on Q9. Please also take the Notes given in the introduction section into account when you have comments.

**Q9: Do companies agree that more RAN1 inputs are still needed to judge whether to consider model monitoring as one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | Model monitoring is still under RAN1 discussions, and there are some per use cases discussions as well. So more RAN1 inputs will be helpful. |
| Xiaomi | Yes | Need more RAN1 input on performance monitoring |
| NEC | Yes | We may need more input from RAN1 before answering the question. |
| vivo | Yes | RAN1 may further downselect the options of model training. |
| ZTE | Yes |  |
| Apple | Yes |  |
| Ericsson | See comment | If model IDs are needed, shouldn’t monitoring be one of the main focuses of RAN2? i.e., can the motivation come from a RAN2 perspective as well? |
| T-Mobile USA | Maybe | I disagree with the rapporteur’s assumption for case 1 and case 2. For case 1 and case 2 how does the MNO map model performance to a particular inference if there’s no unique identifier for the model? If a feature uses 1000’s of models based on hardware, vendor, morphology and the feature perform badly then what? How does a MNO identify the root cause of the problem?  Model ID I required for Case 1-5. |

**Placeholder for summary**:

Regarding to UE capability LCM block, rapporteur believes this LCM purpose is tightly linked to the discussion on model identification and functionality identification. If only functionality identification procedure is introduced and UE capability signaling is used for functionality identification, model ID may be not needed in the UE capability signaling, but it’s still possible to include model ID into UE capability signaling if model identification procedure is considered, so from RAN2 perspective, the safer way is to wait for more progress from RAN1 for this LCM purpose.

Companies are invited to share views on Q10. Please also take the Notes given in the introduction section into account when you have comments.

**Q10: Do companies agree that more RAN1 inputs are still needed to judge whether to consider UE capability as one of the use cases for model ID? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | For UE capability reporting for AIML, we think one reasonable condition should be:  **Both UE and NW has synchronized on model info including model ID**  And then, the UE can consider to include model ID as part of AIML related capabilities. If needed, we think RAN2 can firstly discuss the above condition to see if it is agreeable or not.  In general, we are open to discuss UE capability (including model ID) in RAN2, if most of companies want. However, it seems that UE capability for AIML may face quite a lots of challenges, and model ID is just one of parameters. |
| Xiaomi | Yes | We understand there may be two options to report the AI model related capability,  Option 1: UE reports the supported model (ID).  Option 2: UE reports the AI related capability, e.g. computing capability, storage capability. NW decide which model can be supported by UE according to model’s corresponding requirement.  There are pros and cons for these two options. We can further study. |
| NEC | Yes |  |
| vivo | Yes |  |
| ZTE | See Comments | It depends on whether the ‘global’ model Id can be realized or not, that is , if NW and UE can have a unified understanding on one specific AI model via a global model Id, then the model Id can be used in the UE capability. |
| Apple | Yes (RAN2 wait RAN1 conclusion) | First, we think the solution for UE to report model ID in capability signaling was discussed in RAN1 but not agreed due to diverse opinion:   * Some companies have concern whether it is feasible for NW to manage all possible models via capability signaling. * Only functionality reporting was agreed to use capability signaling as baseline.   Thus, we don't think RAN2 can make conclusion.  Secondly, on the proposal of reporting computing and storage capability, RAN1 is also discussing. As usual business in 3GPP, the lower layer capability is first discussed in RAN1 and then notify the outcome to RAN2 for signaling design. |
| Ericsson | See comment | If we start by assuming that these model IDs are dynamic. Then we can already now agree not to consider model IDs as being part of the UE capability reporting framework.  Then RAN2 needs to analyze whether there is a need to consider an alternative framework to indicate the applicability of these models under more dynamic conditions. |
| T-Mobile USA | No | It’s pretty obvious that UE capability signaling isn’t scalable for the large number of model ID’s, it could be used for Functional ID. |

**Placeholder for summary**:

If any other use cases are considered to be useful for the usage of model ID, please add them into the following table. Please also take the Notes given in the introduction section into account when you have comments.

|  |  |
| --- | --- |
| **Company** | **Comments** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Placeholder for summary**:

### 2.1.2 The type of model ID for identified use case

In RAN2#121 meeting, RAN2 agreed that the model ID should be unique “globally” [1]:

* RAN2 assumes that Model ID is unique “globally”, e.g. in order to manage test certification each retrained version need to be identified.

But it’s still unclear what kind of “global” unique ID we want. At least three directions can be considered based on companies’ inputs [9][18][19].

**Direction1: Pre-defined global unique model ID**

One model ID is assigned to a model algorithm in static manner, i.e. the meaning of each model ID is predefined in the spec like global slice ID, which means all UEs in the same communication system have the same understanding on the meaning of the same global unique model ID no matter which operator the UE has been registered;

**Direction2: dynamically assigned global unique model ID via specific ID management node hosted by a specific operator**

One model ID is assigned to a model algorithm in dynamic manner, i.e. each model ID is assigned via implementation like 5G-GUTI by a specific ID management node hosted by a specific operator, this operator assigned model ID can still be global unique if operator ID info, i.e. PLMN ID, is added as part of the model ID.

**Direction3: dynamically assigned global unique model ID via a specific ID management node shared across operators.**

One model ID is assigned to a model algorithm in dynamic manner, i.e. each model ID is assigned via implementation by a specific ID management node shared across operators, this ID management node is within 3GPP system or out of 3GPP system. The specific ID management node can guarantee that each assigned model ID is global unique.

From offline rapporteur point of view, all three directions can be workable, but the pro and cons for each direction are also quite different if we go deeper. Considering this is the initial discussion to clarify the meaning of global unique ID, companies may need more time to consider the pro and cons for each direction, so it’s not easy to preclude any option in this meeting. More addition, the applied use cases for each direction may be also different, companies may also need to consider the scalability for each direction. To make some progress in baby step, rapporteur suggests to consider the above three directions as the starting point, companies can further study the feasibility for each direction in the future meeting.

Companies are invited to share views on Q11. Please also take the Notes given in the introduction section into account when you have comments.

**Q11: Do companies agree to consider the following global unique model ID definition methods for further study? Please also provide your comments in the comment column if any.**

* **Direction1: Pre-defined global unique model ID.**
* **Direction2: dynamically assigned global unique model ID via specific ID management node hosted by a specific operator.**
* **Direction3: dynamically assigned global unique model ID via a specific ID management node shared across operators.**
* **Please add more if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Direction 1/2/3 | We think all directions are possible, and we may have more analysis/comparisons for them. Other directions are also possible (may be added later). |
| Xiaomi | Yes with comments | Agree with the intention. However, we understand the global unique model ID may be more semi-static rather than dynamic changed, even if it can be assigned by operators. So we suggest to remove the ‘dynamically’ in direction 2/3. |
| NEC | See comments | We would prefer not to say model ID is global at least at this stage. In our understanding, model ID may be a logical ID, which can be UE specific, e.g., assigned by NW during model identification/registration procedure.  Please also note the same discussion is being taken in RAN1, and in RAN1 yesterday GTW session, RAN1 rapporteur of this topic promised to coordinate the parallel discussions. |
| vivo |  | Agree with Xiaomi to remove ‘dynamically’ |
| ZTE | All direction are fine | All directions are possible, and not preclude any further direction.. |
| Apple | All directions | We agree with Xiaomi to remove "dynamically" in direction 2 and 3 because "dynamically" generally implies its time-scale is fast. "Assigned global model ID .." should be sufficient. |
| Ericsson | See comment | In D2 and D3, only referring to “operators” seem to be a bit restricting. So why not simply leaving:  *“dynamically assigned global unique model ID via specific ID management node”* |
| T-Mobile USA | See Comment | It’s not clear what is meant by “model algorithm”, does that refer to a model inference or to a model without inference? We could live with direction 2 and direction 3 if they were reworded as follows:   * Direction 2 : **Each inference is assigned a global unique model ID via specific ID management node hosted by a specific operator.** * **Direction 3: Each inference is assigned global unique model ID via a specific ID management node shared across operators.** |

**Placeholder for summary**:

In [9][14][15][18][19][23], companies have the view that global model ID may have multiple fields and very long length, it’s not efficient to use the global model ID directly for model control. During model control in model-ID-based LCM, an appropriate AI/ML model among a set of AI/ML models is chosen for usage, which can fit the scenario/configuration/site. Considering model activation/deactivation/switch/selection only occurs when UE is in CONNECTED state, a temporary model index can be assigned to each model through model configuration. Model activation/deactivation/switch/selection and fallback can rely on the model index, similar as SCell index for model activation, deactivation, switching and fallback. Based on the model index, both UE and network knows which AI/ML is being in use and monitored.

Based on the views above, companies are invited to share views on Q12. Please also take the Notes given in the introduction section into account when you have comments.

**Q12: From RAN2 perspective, if model-ID-based LCM procedure is introduced, do companies think that we need to consider another type of model ID, i.e. temporary/local model index, for some LCM purposes, e.g. model activation/deactivation/switch/selection? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | FFS | Firstly, we think global model ID can work for model control procedures. The potential issue is about overhead, e.g. it may lead to some overhead if model ID is frequently used in Uu interface. However, we think the overhead issue can be marked and further discussed, and then we can see where to go.  For local model ID or short model ID, we understand the principle is that the original model ID may be long and NW can convert it to a shorter ID to be used in model control. It may save some overhead, but it also needs more thinking on solutions, e.g. the size of temporary/local model ID, how the NW does the mapping, how it works during the mobility cases, and etc. |
| Xiaomi | Yes | The global ID may be long. After model identification, UE and NW have reached consensus on the candidate model. Short/local ID can be mapped to the candidate model, similar as the DRB/SRB ID. During the following LCM, using short/local ID is enough. The overhead of global unique ID may be non-neglectable considering more use cases are to be supported in future. |
| NEC | Yes and comments | We think that global model ID may not be needed (or can be discussed) but local ID is more important. |
| vivo | FFS | Can be discussed in the WI phase as it is a signaling optimization. |
| ZTE | FFS | It depends on the “global” model Id length. And also, we think this is a issue need to be discussed in the WI stage rather than SI stage since we even have no idea about what is the ‘global’ model Id like. |
| Apple | Yes | At least for model selection/activation/deactivation/switching, we think it can be a local ID because these procedures are after model identification (i.e. aligned understanding between NW and UE is achieved). We assume it is configured by gNB and the benefit of a local ID:  1) reduce overhead  2) alleviate some security concern. |
| Ericsson | See comment | As highlighted by Huawei, this depends on how model IDs are shaped (outcome of Q11’s discussion). |
| T-Mobile USA | No | This approach isn’t scalable! Suggest using something similar to the CHO mechanisms. Have a functional ID authorize the use of predefined set of model ID’s. UE has autonomy to pick the appropriate inference for the particular morphology. RAN2 needs to discuss efficient mechanisms to convey model metadata using predefined databases i.e. UE capability ID. |

**Placeholder for summary**:

If anything is missing in this sub-clause and deserves to consider, please add your view into the following table. Please also take the Notes given in the introduction section into account when you have comments.

|  |  |
| --- | --- |
| **Company** | **Comments** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Placeholder for summary**:

## 2.2 Discussion on Model Meta data

### 2.2.1 Use case discussion for Meta data

For Model meta info, there was no much discussion in RAN2 and only a high-level agreement was made for further discussion [3]:

* R2 assumes that from Management or Control point of view mainly some meta info about a model may need to be known, details FFS.

Based on above agreement, it’s obvious that nearly all things are still open for model meta data. Before going to details for model meta data, we should first identify the potential use cases. In other words, for what use cases, model meta data may be beneficial for model management/control.

From offline rapporteur point of view, model meta data includes a set of parameters which can be used to describe different aspects for a specific AI/ML model. The model meta data receiver/user will consider this kind of model description info to handle different LCM purposes.

Based on RAN1 working assumption, rapporteur thinks model identification can be considered as one of the use cases for model meta info [5]:

Working Assumption

|  |  |
| --- | --- |
| Terminology | Description |
| Model identification | A process/method of identifying an AI/ML model for the common understanding between the NW and the UE  Note: The process/method of model identification may or may not be applicable.  Note: Information regarding the AI/ML model may be shared during model identification. |

**Observation2: According to Model identification definition, RAN1 assumes** **information regarding the AI/ML model may be shared during model identification.**

Several proponents also think meta data can be used for model identification [9][13][20]. Although RAN1 has not yet decided what kinds of info is included in the information regarding the AI/ML model, even not clarified the relationship between information regarding the AI/ML model and model meta data. It seems model identification definition given by RAN1 implies that model meta data may be needed during model identification procedure.

Companies are invited to share views on Q13. Please also take the Notes given in the introduction section into account when you have comments.

**Q13: From RAN2 perspective, if model identification procedure is introduced, do companies think that model identification procedure is one of the use cases for model meta data? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes | It is reasonable to consider meta data for model identification. In our paper [20], we have the following analysis:   * **5 Model registration (or model identification).** As we analysed above, for model identification, it may happen that the UE just gets the model information (e.g. via Solution 4). Then, the UE may need to register the model to 3GPP network, so the UE may need to include the meta-data during the model registration. For the content, we have the similar views as for model transfer/delivery. |
| Xiaomi | Yes with comments | We understand the NW and UE shall have common understanding about model’s meta info. But it does not mandate UE and NW shall always exchange meta info. For example, if a model is predefined in spec or provided by application server, meta info of this model is available at UE and NW already. |
| NEC | Yes |  |
| vivo | Yes | The meta-data can be exchanged during model identification, e.g., model meta-data is transferred along with model ID. |
| ZTE | Comments | It depends on how the global model Id is obtained by UE.  If the global model Id is authorized by the NW via some kind of procedure (i.e. model registration), it means the NW and UE have a very much unified understanding about the model indicated by the global Id, in this assumption, we do not see any need to introduce the meta data info since all the meta information has been known by the NW/UE during the registration procedure.  If the global model Id is authorized by the third party or the model Id is not absolutely ‘global’ for the NW, that means the meat info maybe needed since the NW may need some additional information that is not included in the model Id. |
| Apple | Yes | We agree with Xiami that meta info is optional information to describe the model in model identification, but model ID should be mandate.  Again, we disagree to use terminology "model registration", which is not agreed by RAN1. RAN2 should only use terminology of "model identification", to align with RAN1. |
| Ericsson | Yes, see comment | We think this is already discussed above (see Observation 1 before Q1). |
| T-Mobile USA | Yes | There are ways to convey large amount of meta data without transferring the information across the network i.e. prepopulating databases during the model ID registration process. |

**Placeholder for summary**:

Network triggered LCM purposes, e.g. model selection/activation/deactivation/switching/fallback, get wider support based on RAN1 contributions, but how the network triggers the corresponding LCM action may rely on UE assistant info and/or local stored model meta data. Network can use the model meta data as one of the inputs to make decision for subsequent model control/management.

Companies are invited to share views on Q14. Please also take the Notes given in the introduction section into account when you have comments.

**Q14: From RAN2 perspective, do companies agree that model meta data can be used for the following use cases for model control purpose: model selection/activation/deactivation/switching/fallback? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | FFS | In our understanding, if the model has been synchronized between UE and NW (e.g. via model identification), model ID is the main info to do the model control, while the usage of other info during model control should be clarified.  In addition, the question is not very clear, it may lead to some understandings, e.g.:   * Whether model meta data can be used for UE or NW decision * Or, whether model meta data can be used for the signalling related to model control   For the 1st bullet, we think it should be possible and it can be left to implementation. We hope that the offline rapporteur can clarify the question. |
| Xiaomi | Yes | This question is related to the content of the meta info. We understand the meta info may include model’s applicable use case and scenario. Therefore, if the use case or scenario changes, NW/UE can perform LCM, e.g. activation/deactivation/switching/selection, accordingly. |
| NEC | FFS | Same understanding as Huawei. It is unclear if the question is to ask the UE-NW coordination or the network internal decision (on how use the meta data) which is implementation specific. |
| vivo | Yes | For the UE side model, the UE may select/activate/deactivate the model based on the applicable condition in the meta-data. |
| ZTE | See comments in Q13 |  |
| Apple | FFS | Same view as Huawei. |
| Ericsson | See comment | It appears that Huawei is already hinting this in their answer.  Couldn’t we already focus on the ID itself?  In principle, disentangling model ID from meta data could end up being representing a huge burden, as e.g., the NW would need to understand and take decisions considering both separately. |
| T-Mobile USA | Yes |  |

**Placeholder for summary**:

Although how the network makes the decision based on meta data is usually up to implementation, it’s still needed for network to get valid model meta data via offline manner or 3GPP visible signaling before subsequent model control operation.

Companies are invited to share views on Q15. Please also take the Notes given in the introduction section into account when you have comments.

**Q15: From RAN2 perspective, do companies agree that, for network-controlled model, i.e. UE side model/network side model/two side model, network may need to get valid model meta data via offline manner or 3GPP visible signaling in advance before subsequent model control operations? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes, but | Firstly, we prefer to remove “**network-controlled model**” and to directly use “**UE side model/network side model/two side model**” instead, in order to avoid confusions.  Secondly, in our paper [20], we think online/offline model identification are two possible ways, which can be used as inputs for more RAN2 discussions. |
| Xiaomi | Comments | The ‘network-controlled model’ may be confusing. Why the UE-controlled model is excluded? In general, we understand the node which makes LCM decision should be aware of the model’s meta info. The node can be UE or network. |
| NEC | Yes with comments | offline manner or 3GPP visible signaling can changed to 3GPP signaling or UE-Network interaction transparent to 3GPP signaling |
| vivo | Comments | Whether the network needs the model meta-data is based on the functionality mapping. E.g., for the UE-sided model trained by the OTT server, the NW may be agnostic about the meta-data. |
| ZTE | Yes | The suggestion from HW is fine to us. |
| Apple | Yes with comments | We also think “**network-controlled model**” can be removed. Meanwhile, the meta information should be optional, so we should keep "may". |
| Ericsson | See comment | This will depend on whether the meta data is part of the ID itself, or? If so, an ID could in principle be enough. |
| T-Mobile USA | Yes | See earlier comment |

**Placeholder for summary**:

In last RAN2 meeting [1], RAN2 agreed to further analyze the pros and cons for each solution on model transfer/delivery, but RAN2 has not yet discussed what kinds of data should be transferred/delivered. The initial consideration is that at least model algorithm data which includes model structure and model weight parameters will be transmitted during model transfer/delivery procedure. But model algorithm data is not enough as the UE still doesn’t know what functionality this model algorithm data is used for and other essential model description parameters which is necessary for model usage.

Several proponents also think meta data can be used for model transfer/delivery [9][20]. From offline rapporteur perspective, other model description parameters may still be needed if UE wants to use the AI model after model transfer/delivery. For example, model input/output info, model version info, model format info, model accuracy info and so on. These model meta info may be essential for model usage.

Based on the views above, companies are invited to share views on Q16. Please also take the Notes given in the introduction section into account when you have comments.

**Q16: From RAN2 perspective, do companies think that model transfer/delivery procedure is one of the use cases for model meta data? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Huawei, HiSilicon | Yes |  |
| Xiaomi | Yes | Meta info can be indicated together with the actual model during model delivery. |
| NEC | Yes |  |
| vivo | Yes | Meta-data can be exchanged along with the model ID during model transfer. |
| ZTE | See the comments in Q13 |  |
| Apple | Yes |  |
| Ericsson | Yes, but… | Again, we need to understand whether the meta data is part of the ID or not. |
| T-Mobile USA | No | Model ID is the index for all model meta data then maybe no need to transfer model data. |

**Placeholder for summary**:

If anything is missing in this sub-clause and deserves to consider, please add your view into the following table. Please also take the Notes given in the introduction section into account when you have comments.

|  |  |
| --- | --- |
| **Company** | **Comments** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Placeholder for summary**:

### 2.2.2 Content discussion for Meta data

During Week1 online discussion, RAN2 made the following agreement for meta data:

* R2 assumes that Information such as FFS:vendor info, applicable conditions, model performance indicators, etc. may be required for model management and control, and should, as a starting point, be part of meta information.

Based on contributions submitted to this meeting [9][11][12][13][14][16][17][19], rapporteur thinks the following kinds of meta data can be further considered:

**a, model input info;**

**b, model output info;**

**c, model version info;**

**d, model format info;**

**e, required AI capability;**

**f, vendor info;**

**g, applicable scenario, configuration, site information;**

h, computational complexity: FLOPs, level of pre-/post-processing;

i, model complexity: number of real-value model parameter, number of real-value operations;

**j, model size;**

k, model performance: Model accuracy, model bias, model variance;

**l, model functionality;**

Companies are invited to share views on Q17. Please also take the Notes given in the introduction section into account when you have comments.

**Q17: From RAN2 perspective, which kind of meta data can be further considered? Please also provide your comments in the comment column if any.**

|  |  |  |
| --- | --- | --- |
| **Company** | **a~l (more than one can be selected)** | **Comments** |
| Huawei, HiSilicon | OK: a, b  **Add use cases and types** | In our paper [20], we propose the following two info for meta info, which can be also discussed here.   * + **What (sub) use cases are applicable for a specific model, e.g. CSI/BM/Positioning**   + **What types are applicable for a specific model, e.g. UE-sided model/UE part model for two-sided model**   **We have some concerns/comments about the following info:**  For **d, model format info**, we think more RAN1 inputs will be helpful, as RAN1 is still discussing proprietary/open format and some model transfer options.  For **e, required AI capability**, **h, computational complexity: FLOPs, level of pre-/post-processing**, and **j, model size**, we think they are related to UE capability discussions, so some UE capability progress (from RAN1 and RAN2) will be helpful.  For **f, vendor info**, we would like to understand more about the motivation, usage and whether there are user privacy issues.  For **g, applicable scenario, configuration, site information**, we think “site information” is sensitive and the motivation/usage should be clarified.  For **k, model performance: Model accuracy, model bias, model variance**, we wonder about the usage, e.g. whether it is to be used for testing purpose or not.  For others, we are open. |
| Xiaomi | E, g, h, J,I | We understand at least meta info shall include the applicable scenario (g) and the required AI capability (e, h, I, j).  A, b, we think this can be implicitly indicated by the applicable scenario.  C, it’s unclear whether version is supported.  D, depends on whether model format is supported.  F, we don’t see the need of vendor info.  K, the model performance, e.g. accuracy, may change due to environment change. We assume meta info should be static info.  L, the relation between model and functionality is unclear. |
| NEC | A/B/C/E/J/L/G | In addition, we would like to add “model monitoring method” into “meta data”. |
| vivo | g,  (sub)Use case | Revise data g to ‘applicable condition’  a/b, model input and output can be specified for each use case.  c, model version info can be implicitly indicated by unique model ID.  d, model format info needs RAN1 further input.  e, required AI capability, the NW will get the UE capability beforehand and will only transfer the model that can UE support to the UE.  f, vendor info can be implicitly indicated by a unique model ID.  g, applicable condition is needed for further LCM purposes.  h/i, model complexity, similar to the required AI capability, the NW will get the UE capability beforehand and will only transfer the model that can UE support to the UE.  j, model size, the UE will know the model size when it receives the model.  k, model performance, is not necessary as UE will monitor the performance when needed.  l, model functionality needs RAN1 further input on the definition of functionality.  Agree with Huawei that (sub)use case can also be indicated as meta data. |
| ZTE | Comments | Too early to discuss, to our understanding, what the meta information is tightly related to the what contents including in the ‘global’ model Id also related how the model Id is obtained by UE.  Before discussing this, we need make some kind of the working assumptions about the model Id fields, for example, assuming that the model Id is on dimension-Id, what kind of meta information is needed for different LCM purposes. |
| Apple | a, b, d, j, g (but only **applicable scenario)** | First, we think below are needed   * **a, b**   + They are important information of model description. We don't think it can be implicitly derived by scenario (e.g. size of input) * **d, model format info** * **j: model size**   + It is necessary for the UE to determine whether to download the model based on its remaining memory.   We have concerns on below options:   * **c, model version info**   + Based on RAN2 agreement " Model ID is unique “globally”, e.g. in order to manage test certification each retrained version need to be identified ", we think model ID can identify model version * **f, vendor info**   + We think the global unique model ID can already identify the vendor. So, the further vendor information is not needed.   + We think further vendor info is sensitive and there are user privacy issues.   + We don't understand why 3GPP need to specify vendor information. * **e, h, i**   + 2e think they are related to UE capability discussions. And RAN2 should wait RAN1. * **g**   + We think only **applicable scenario is valid. Don't see need of configuration and site information** * **k**   + We also think it is dynamic information, which seems not suitable to include in meta info * L   + Agree with Xiaomi that the relation between model and functionality is unclear. * Use case   + We think model ID can identify it (at least for CSI, BM and positioning) |
| Ericsson | See comment | Motivations to having these should come from RAN1 analysis. It appears difficult to see how this can end up in a constructive agreement in RAN2.   We can then stick to the assumption we have right now (see agreement above). In fact, as we see that the discussion is perhaps shifting to less RAN2-centric matters, why not agreeing to the following:  *“R2 assumes that model IDs allow for model management and control. Depending on RAN1 input, RAN2 can later discuss whether there is a need to consider additional signaling/procedures conveying extra meta data”* |
| T-Mobile USA | See comment | All of the above is applicable, we would add Morphology and a geographic information. Until models can apply to many use cases it will be necessary to understand what they were optimized for. |

**Placeholder for summary**:

If anything is missing in this sub-clause and deserves to consider, please add your view into the following table. Please also take the Notes given in the introduction section into account when you have comments.

|  |  |
| --- | --- |
| **Company** | **Comments** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Placeholder for summary**:

# Conclusion

the summary proposals are listed as below:

**Easy agreements if any:**

**Potential agreements:**

**Open Issues (which seem important to address):**

# References

[1] 3GPP RAN2#121 meeting Chairman Notes.

[2] 3GPP RAN2#120 meeting Chairman Notes.

[3] 3GPP RAN2#119bis meeting Chairman Notes.

[4] 3GPP RAN1#110bis meeting Chairman Notes

[5] 3GPP RAN1#111 meeting Chairman Notes

[6] 3GPP RAN1#110 meeting Chairman Notes

[7] 3GPP RAN1#112 meeting Chairman Notes

[8] 3GPP RAN1#109 meeting Chairman Notes

[9] 3GPP R[2-2302546](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2302546.zip) Discussion on Model ID and Model Meta Data OPPO discussion Rel-18 FS\_NR\_AIML\_air

[10] 3GPP R[2-2302649](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2302649.zip) AIML architecture Nokia, Nokia Shanghai Bell, T-Mobile US discussion Rel-18 FS\_NR\_AIML\_air

[11] 3GPP R[2-2302746](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2302746.zip) General architecture assumptions, model ID and entity mapping Intel Corporation discussion Rel-18 FS\_NR\_AIML\_air

[12] 3GPP R[2-2302953](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2302953.zip) Discussion on Architecture General vivo discussion Rel-18 FS\_NR\_AIML\_air

[13] 3GPP R[2-2303017](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303017.zip) Further discussions on architecture general aspects of AIML for NR air-interface CATT, Turkcell discussion Rel-18 FS\_NR\_AIML\_air

[14] 3GPP R[2-2303122](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303122.zip) Discussion on architecture aspects Xiaomi discussion

[15] 3GPP R[2-2303371](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303371.zip) Discussion on AI/ML model identification, LCM and capability Apple discussion Rel-18 FS\_NR\_AIML\_air

[16] 3GPP R[2-2303580](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303580.zip) Discussion on general AI architecture Spreadtrum Communications discussion Rel-18

[17] 3GPP R[2-2303674](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303674.zip) Discussion on AI/ML Architecture General Qualcomm Incorporated discussion Rel-18

[18] 3GPP R[2-2303760](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303760.zip) Model ID and Mapping of Functions to Physical Entities MediaTek Inc. discussion

[19] 3GPP R[2-2303885](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303885.zip) Discussion on AI/ML model identification and functionality identification Futurewei Technologies discussion

[20] 3GPP R[2-2303893](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303893.zip) Discussion on model ID and mapping of functionality to entities Huawei, HiSilicon discussion Rel-18 FS\_NR\_AIML\_air

[21] 3GPP R[2-2303946](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2303946.zip) Model identification and LCM aspects of AI/ML for NR air interface AT&T discussion

[22] 3GPP R[2-2304116](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2304116.zip) Architecture and management for AIML Ericsson discussion Rel-18 FS\_NR\_AIML\_air

[23] 3GPP R[2-2304173](file:///E:\3GPP文档\会议文稿\2023\RAN2%20121b\R2-2304173.zip) AIML method\_Architecture General LG Electronics discussion Rel-18