3GPP TSG-RAN WG2 #123 R2-23XXXXX

Toulouse, France, August 21 – 25, 2023

Agenda Item: X.Y.Z

Source: Ericsson

Title: Outcome of [Post122][059][AIML]

Document for: Discussion

# 1 Introduction

The scope of the discussion is given by the following email thread:

|  |
| --- |
| * [Post122][059][AIML] TR text proposal (Ericsson)   Scope: Assemble agreed figure, tables etc into a TR baseline TP. Identify discussion points that seems essential to progress RAN2 TP in the near term  Intended outcome: Agreeable TP,  Deadline: Long |

The main purpose of this document is to address the discussion related to the functional framework, specifically regarding the data/information flows (i.e., arrows) which we did not have time to discuss online.

If you have not already done so, please refer to the separate document (i.e., RAN2’s TP to the TR) before proceeding with this document. Please note that an Annex has been included in that document listing all RAN2 agreements. The Rapporteur has highlighted those that are captured as TP.

Additionally, please provide any further details you believe are worth addressing in this email discussion as input to Q2 (see below).

**Deadline for comments: Thursday Aug 10th, 2023, 1000 UTC**

**Inactive periods and other planning comments:**

July 1st – 30th 3GPP Inactive Period

August 10 1000 UTC Deadline Long Email Discussions

August 11 1000 UTC Submission Deadline RAN2#123

Below you can find the list of participating companies and their respective responsible delegates.

|  |  |  |
| --- | --- | --- |
| **Company** | **Delegate name** | **Email address** |
| Ericsson | Felipe Arraño Scharager | felipe.arrano.scharager@ericsson.com |
| vivo | Boubacar Kimba | kimba@vivo.com |
| Xiaomi | Xing Yang | Yangxing1@xiaomi.com |
| Lenovo | Congchi Zhang | Zhangcc16@lenovo.com |
| Qualcomm | Rajeev Kumar | rkum@qti.qualcomm.com |
| Apple | Peng Cheng | pcheng24@apple.com |
| CATT | Da Wang | wangda@catt.cn |
| Huawei, HiSilicon | Jun Chen | [jun.chen@huawei.com](mailto:jun.chen@huawei.com) |
| LGE | Soo Kim | [soo.kim@lge.com](mailto:soo.kim@lge.com) |
| Spreadtrum | Xiaoyu Chen | [xiaoyu.chen@unisoc.com](mailto:xiaoyu.chen@unisoc.com) |
| Interdigital | Oumer Teyeb | [Oumer.teyeb@interdigital.com](mailto:Oumer.teyeb@interdigital.com) |
| ZTE | Fei dong | Dong.fei@zte.com.cn |

# 2 Discussion

## 2.1 On the functional framework

The following was agreed during RAN2#122:

|  |
| --- |
| Functional Arch   * Intention is to cover functional arch in general, e.g. covering both be model based and/or functionality based LCM * “Model Storage” in the figure is only intended as a reference point (if any) for protocol terminations etc for model transfer/delivery etc. It is not intended to limit where models are actually stored. Add a note for this. * Remove “Model” in Model Managemt and Model Inference and for the actions/the arrow form Management to Inference (to reduce the risk for misunderstanding). * Management may be model based management, or functionality based management. Add a mote for this. * With the modifications above Figure 2 from R2-2305327 is agreed |

Figure 2 from [R2-23053207](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_122/Docs//R2-2305327.zip) can be seen below:

|  |
| --- |
| **Figure 2: Functional architecture of AI for air interface** |

As per the agreements above, the Rapporteur has captured the following Figure in RAN2’s (currently discussed) text proposal for TR 38.843:

|  |
| --- |
| Figure 4.4-1: Functional framework for AI/ML for NR Air Interface |

Please be aware of the following changes to the names of the data/information flows (i.e., arrows):

* The arrow from Management to Inference has been renamed as “Management Instruction”,
* The arrow from Management to Model Training has been renamed “Monitoring output”,
* The arrow from Inference to Management has been renamed to “Inference Output”.

Arguments to these changes are provided as input to Q1 below.

**Question 1:**

Companies are invited to provide their views on the functional framework in the Figure above (Figure 4.4.-1). Try focusing on the data/information flows (i.e., the arrows), as the main blocks/functions have already been discussed and agreed online. Hence, try answering: should an arrow be optional? Should we rename them? Should we add/remove some? etc...

*Note: As per RAN2’s agreement the Rapporteur’s intention is to update the TR’s Figure that will be submitted to RAN2#123 according to the outcome of the discussion below.*

|  |  |
| --- | --- |
| **Company** | **Views** |
| Ericsson | **On the arrow’s names:** We think that we should rename the following arrows as follows:   *Model selection/(de)activation/switching/fallback 🡪 Management Instruction  Monitoring output 🡪 Inference output Performance Feedback / Retraining Request 🡪 Monitoring output*  For the first one, there is a need to generalize what this is intended for.  Further, on the other ones, please look that monitoring data goes into management, then the outcome of management should be “monitoring output”, and the outcome of inference should be “inference output”.   Note that for the “Performance Feedback / Retraining Request”, the Model Training function should be the one to decide whether retraining is needed based on the monitoring output (i.e., not the Management function deciding this).   **On the presence/optionality of arrows and functions (blocks):** *- (arrow) Model Transfer/Delivery Request:* We are not entirely sure why the management function should be in charge of this. In principle, the management entity should simply provide the outcomes of certain model performances. Hence, we are inclined not to proceed with this model (update) request.  Additionally, since we agreed that “Model Storage” is only intended as a reference point, we could also use dashed lines for all arrows from/to the “Model Storage” block (eventually the block itself could also be dashed). |
| vivo | First:  We think we should decouple the TR update and whether to rename the arrows in the figure. That is, the TR update should just focus on the current RAN2 agreements. And separately we can discuss whether to rename the arrows. If RAN2 agree to do the update, we still have RAN2#123 post meeting for TR update. There is no rash to do this in the current TR update discussion.  Second:  On:***Model selection/(de)activation/switching/fallback  Management Instruction***:We think using “Instruction” may be confusing. We do not see the necessity to introduce a new terminology for current model management procedure related to *selection/(de)activation/switching/fallback .*  *On* ***Performance Feedback / Retraining Request  Monitoring output***  *Retraining Request* can be based on monitoring output, but not every monitoring output would require retraining request. So, whether or not there is need to retrain a model should be decided at management entity. If we agree to rename, as proposed by Rapp, that means the model retraining decision would be totally left to model training entity. That is not RAN2 current agreement. |
| Xiaomi | **On the arrow name**  The functionality of the arrows should be aligned among all the arrows in the diagram. Now it seems that the arrows have different meaning in the diagram. Some arrows represent the contents to be exchanged between two blocks. For example, the arrow between data collection and model training/ management/inference. While, some other arrows represent involved process between blocks. For example, the arrow between management and model inference. In our view, it is more general to make one arrow represent involved procedure. Considering this aspect, the text of arrow between data collection and model training/management/model inference and the text of arrow between model training and model storage can be updated to reflect involved procedure.  **On the arrow addition**  The functionality/model identification process is one important aspect in the whole LCM. This procedure may be reflected in the diagram. Management is responsible to manage the model or functionality identification. After the model is trained, then related functionality identification or model identification can be performed if needed. Hence, an arrow representing the model/functionality identification can be added between the model training and management.  **On the arrow removal**  The management need to acquire performance monitoring inputs. In the fig, there are two arrows from data collection and inference. However, we understand the data collection shall support both legacy metrics and AI related metrics. Therefore, seems the arrow from model inference to management is unnecessary and can be removed. It is already covered by the arrow between data collection and management since the collected monitoring data could also include the output of the inference.  **On the optionality of arrow**  We are not sure whether the arrow from management to model training is always mandatory. Seems the only use case of such arrow is to retrain the model. However, we understand the model training may be out of 3GPP, which means the training may be independent from the performance monitoring. Therefore, we suggest to make this arrow optional.  Also agree with Ericsson, the model storage related arrow can be optional since the model storage itself is optional. |
| Lenovo | **On the arrow’s names:**  *Model selection/(de)activation/switching/fallback 🡪 Management Instruction  Monitoring output 🡪 Inference output Performance Feedback / Retraining Request 🡪 Monitoring output*  Regarding the proposal from rapporteour, we feel generalizing the terminology *Management Instruction, Monitoring output* may not help with the interpretation and readability of the framework. The original wording reflects the intention better.  *Monitoring output 🡪 Inference output* is a good suggestion, we understand it is upon the Management function that does the monitoring, and it may take the inference output (e.g. the predicted CSI/BM result) into account.  **On the presence/optionality of arrows and functions (blocks):**  We are ok to make the Model Transfer/Delivery Request optional as Rapporteur suggested.  In addition, maybe the *Inference output* after change could be optional too. The monitoring could be upon the overall performance rather than comparing the predicted result with the ground truth. |
| Qualcomm | I think we should try to make only required changes.  On Arrow naming:  -----------------------------  Renaming monitoring output to inference output: Agree.  Renaming performance feedback / retraining request to Monitoring output: Not required. Management functional block can make provide “performance feedback” to training entity and final decision is taken as the training entity. Or, retraining decision can be taken by entity implementing management functional block and retraining request can be sent to the training entity.  Renaming Model selection/(de)activation/switching/fallback  Management Instruction to Management instruction: We can simply remove the Model and that should be okay. As, VIVO as mentioned, management instruction may be confusing. |
| Apple | First, we tend to think we can focus on just capturing RAN2 agreements in TR for now. And because this figure is just for illustrated purpose, we prefer to keep the wording of online agreement as much as possible.  Then, we provide our view on above proposals:  **On the arrow rename**  Renaming monitoring output to inference output:  Agree. The previous wording is indeed misleading.  Renaming performance feedback / retraining request to Monitoring output: Not required. We think previous wording is sufficient, although not very generic. Again, because this figure is just for illustration purpose in TR, we think it is sufficient.  Renaming Model selection/(de)activation/switching/fallback  Management Instruction to Management instruction:  Not required. We think previous wording is sufficient, although not very generic. Again, because this figure is just for illustration purpose in TR, we think it is sufficient.  **On the optionality of arrow**  We don't think RAN1/RAN2 is ready to discuss whether one step (one arrow) is optional or mandatory, which is typically discussed in WI or even late stage of WI. Thus, we suggest:  1) No need to further discuss optional vs mandatory (i.e. solid line vs dash line).  2) Add one NOTE on the figure: "The figure is intended to illustrate basic principle of functional framework. It doesn't intend to specify whether any procedure indicated by the arrow is mandatory or optional."  3) Following online agreement spirit, we agree with Rapporteur that the “Model Storage” block can be dashed. But as mentioned in 1), suggest not to discuss whether any line is dashed or solid, which we don't think clear conclusion can be made for now.  **On the arrow addition/remove**  We tend to agree with Ericsson that it is not clear why Model Transfer/Delivery Request is needed between "management" and "model storage". And it is questioned how a "management" block can directly communicate to a NW entity to store model. So, if no valid justification, we suggest to remove this arrow. |
| CATT | We agree with the following two arrow renaming:   * *Model selection/(de)activation/switching/fallback 🡪 Management Instruction* * *Monitoring output 🡪 Inference output*   But for the second one, i.e.,“The arrow from Management to Model Training has been renamed “Monitoring output””, we think it is the management entity’s responsibility to decide whether to perform re-training/fine-tuning. The model training only perform the action of “training” based on the data and/or the feedback/(re-)training request. So “*Performance Feedback / Retraining Request*” in the legacy framework figure seems appropriate.  And we agree with Ericsson that the “Model Storage” related arrows should all be set to dashed lines with other blocks. |
| Huawei, HiSilicon | We are ok with the following change:  *Model selection/(de)activation/switching/fallback 🡪 Management Instruction*  For the arrow “monitoring output”, we wonder whether it is really needed. We have already had “monitoring data” from Data collection function, which is related to monitoring and it is the same as “monitoring output”. **So we suggest to remove the arrow “monitoring output”.**  For the following change, we agree with the intention, but **we suggest to use “Management instruction”**, because this arrow is to provide some instruction for the training and we do not have to list detailed information here.  *Performance Feedback / Retraining Request 🡪 Monitoring output*  On *(arrow) Model Transfer/Delivery Request*, we are not clear about the usage. There is an arrow from model training function to model storage function, which is sufficient. **So we suggest to remove this arrow.**  On model storage and related arrows, we share similar views are Ericsson, and **we suggest to either remove “model storage + related arrows” or use dashed lines/dashed block.** |
| LGE | We agree with the following arrow renaming:   * *Monitoring output 🡪 Inference output*   For others, we think the original naming is sufficient and more readable. |
| Spreadtrum | **On the arrow’s names:**  **For the arrow from Management to Inference, we’d like to keep the original name.** It is clearer and we don’t need to explain the meaning of “model instruction” anymore.  **For the arrow from Management to Model Training, we’d like to change the name to (Re)Training Request.**  The name of Monitoring Output is too general and it may also include the content of the arrow from Management to Inference.  And what Model Training functionality can do is to perform model training. Thus it needs the data for (re)training and the command whether to start (re)training. As the data has been collected by Data Collection functionality and provided in Training data to Model training function. The arrow from Management to Model Training should be **(Re)Training Request.**  **On the arrow addition/removal:**  For **the arrow from Inference to Management**, we think the Inference Output has already been collected by Data Collection functionality, thus this arrow **may not be necessary**.  **On the presence/optionality of arrows and functions (blocks):**  **For Model Storage function**, actually as a logical function, it can physically locate at an independent node or it can physically locate at the node who also holds Management, Model Training or Inference function.  When Model Storage doesn’t collocate with Inference, Model Transfer/Delivery is needed. Otherwise, the Model Transfer/Delivery is not needed.  Therefore, Model Storage is optional. It only exists when not collocated with Inference function. And **the block diagram of Model Storage and the arrows terminated at/from Model Storage should be optional**. |
| Interdigital | **-** In the figure, one thing that is not clear is on how data is collected (no input arrow to the data collection box) and what triggers the data collection part to send the data for model training or performance monitoring/management or inference parts.  We suggest considering having arrows from the management/model training to the data collection (e.g., requesting the data for training, monitoring) (or an alternative will be making the lines have double arrows).  Also, it should be clarified/noted that the figure is assuming offline training, because otherwise (i.e., online training), there may be a need for a direct link/arrow between inference and model training. |
| ZTE | **Agree with the following change:**  *Monitoring output 🡪 Inference output*  **Have no strong point of view on other changes.** |

[Rapporteur’s Summary]: *To be added…*

1. According to companies’ inputs...
2. RAN2 to …

**Question 2:**

Companies are invited to provide additional comments with respect to RAN2’s TP for the TR. Is there something missing? Is there any other topic that should be addressed by this email discussion? etc…

As per what the Chair has described as scope for this email discussion, companies are also invited to provide views on discussion points that appear essential to progress RAN2’s TP in the near term. Please note that the Rapporteur has already added Editor’s Notes to the TR (highlighting some topics that need further study or/discussion/progress in RAN2).

|  |  |
| --- | --- |
| **Company** | **Views** |
| Ericsson | We have tried to stress some points in the TP which RAN2 should start discussing in the near term, these include:   * (Prioritize discussion on) Mechanisms to report updates on applicability of models/functionalities * UE capability reporting considerations (note that RAN1 have agreed some things on this matter, and we believe that RAN2 should have a saying as well). |
| vivo | For 4.2, the EN about functionality and capability sees not needed as further discussion and clarification will be captured in sections 7.3.1.1 and 7.3.1.4.  For 7.3, all the descriptions related to the assumptions in the LS to RAN1 should be rephrased as ‘RAN2 assumes…’. Besides, suggest adding an EN for the RRC state of data collection:  *Editor’s note: Analysis and potential enhancement of the data collection when UE in the non-connected state can be revisited when needed.*  For 7.3.2/3/4 “…be initiated by either the …” We think “decided” or “decided/initiated” may be more precise.  For 7.3.4, the original assumption is 'For model monitoring at NW side, performance metrics can be generated by UE/gNB and terminated at LMF.‘ However, the current TP added the restriction ' For monitoring at the network side of UE-sided model‘. We suppose the gNB is not able to generate performance metrics for UE-sided model, and suggest removing ' of UE-sided model‘. |
| Apple | We echo Ericsson raised two points. |
| Huawei, HiSilicon | For the email scope “Identify discussion points that seems essential to progress RAN2 TP in the near term”, our understanding is that the discussion points should be closely related to TP, but not every open issue. Otherwise, this part will become very broad.  We have checked RAN1 progress, but we fail to find RAN1 agreements on UE capaiblity reporting. Regarding the 1st point, the intention and use cases are unclear to us.  **In summary, we think the two points could be discussed at the coming RAN2 meeting, but not in this email discussion.**  For the RAN2 TP to TR 38.843, we have a general comment as below:  We observe that RAN1 identified functionality-based LCM and model-based LCM, and made some progress. In the past RAN2 meetings, we also used the same terminologies.  For now, there seems common discussions for both types in RAN1/RAN2. From TR point of view, we think that companies may need to review the two types separately in order to understand how each type works. For example, for functionality-based LCM for a specific use case, what are the requirements on some LCM components, what are the solutions, what are possible spec impacts, and etc.  **In this case, we suggest to have separate descriptions for functionality-based LCM and model-based LCM (per LCM componenet per use case) in the TR 38.843. At least, if we have to describe both in one section, there should be separate paragraphs.** |
| LGE | We share a similar view with Vivo for 7.3.2/3/4 in changing "initialized" to "decided".  Additionally, for 7.3.2, in the monitoring description, “UE” can be added. As per the RAN1 agreement, the UE needs to report a performance metric for UE-sided monitoring as well.  *For monitoring at the network/UE side of UE-sided model, the UE can generate performance metrics while the termination point for these metrics is the gNB.* |
| Interdigital | We agree with the two points identified by Ericsson. |
| ZTE | We echo Ericsson’s pointing out. the discussion in RAN2 and the outcome from discussion can give some high level guidance in the following WI but not go too detail. |
|  |  |
|  |  |
|  |  |

1. According to companies’ inputs...
2. RAN2 to …

[Rapporteur’s Summary]: *To be added…*

# 3 Conclusion

In the previous sections we made the following observations:

[Observation 1 According to companies’ inputs...](#_Toc138684978)

[Observation 2 According to companies’ inputs...](#_Toc138684979)

Based on the discussion in the previous sections we propose the following:

[Proposal 1 RAN2 to …](#_Toc138684980)

[Proposal 2 RAN2 to …](#_Toc138684981)

# 4 References

1. R2-23XXXXX, “R2 input to TR 38.343”, Ericsson, RAN2#123, August 2023, Toulouse, France