**3GPP TSG-RAN WG3 Meeting #111electronic R3-210xxx**

**25 January – 4 February 2021**

**Agenda Item: 30**

**Source: CMCC**

**Title: Summary of offline discussion on EnhDataColl**

**Document for: Discussion and Decision**

# 1 Introduction

**CB: # 17\_Basket\_AI**

**- resolution of FFSs**

**- high-level framework for continuation of SI**

**- new use case: energy saving? (lower prio discussion?)**

**- If consensus, capture any general principles / descriptions / open issues / WA / agreements for upcoming RAN3 work (to be captured in Chair’s Notes)**

**- suggest to avoid discussing details (CRs,TPs, etc.) unless there is full agreement**

(CMCC - moderator)

Summary of offline disc R3-210980

We intend to achieve some high-level agreements during the first phase discussion and proceed with the TPs if possible in the second phase.

# 2 For the Chairman’s Notes

**To be added after email discussion.**

# 3 Discussion 2nd Round

The following proposals are proposed for agreement after 1st discussion:

**Proposal 1. Agree the TR 37.817 v0.1.0**

**Proposal 2: AI framework is described from processing point of view.**

**Proposal 3: Change “Data collection” to “Data collection & preparation”, and change “Model training” to “Model training (offline/online)”.**

**Proposal 4: Feedback from action to data sources is performance feedback.**

**Proposal 5: The** **direct** **feedback from action to model training is** **FFS.**

**Proposal 6: Postpone the discussion on other open issues proposed by R3-210617.**

**Proposal 7: The use cases agreed to start from at RAN3#110 E-meeting could be prioritized.**

**Proposal 8: Postpone the discussion on detailed description of use case to next meeting.**

**Question 1: If companies have any further comments on these proposals, please indicate in the below table:**

|  |  |
| --- | --- |
| **Company** | **Comments/Suggestions** |
| **Ericsson** | **Proposal 1. Agree the TR 37.817 v0.1.0**  [AC] OK  **Proposal 2: AI framework is described from processing point of view.**  [AC] Here We should re-formulate and state that “**AI framework is described from a functional point of view”**  **[Moderator] The wording “Processing point of view” may be misleading, it refers to the alternative 2, It seems Ericsson also prefer alternative 2 with modification, so we can focus on the figure itself.**  **Proposal 3: Change “Data collection” to “Data collection & preparation”, and change “Model training” to “Model training (offline/online)”.**  [AC] We would like to maintain “Data Sources” as mentioned in the SoD. We would like to leave the discussion on online/offline model training for the next meeting as this has implications on the functional model.  **Proposal 4: Feedback from action to data sources is performance feedback.**  [AC] we propose to remove this proposal as it anyhow does not change what is captured in the TR. We commented that we would like to work on the description of the data sources instead of having harrows for each type of input information  **Proposal 5: Feedback from action to model training is needed.**  [AC] As mentioned in the SoD, we do not think this is needed and it is already covered by the fact that Data Sources can provide information to model training, one type of such information is performance feedback  **Proposal 6: Postpone the discussion on other open issues proposed by R3-210617.**  **Proposal 7: The use cases agreed to start from at RAN3#110 E-meeting could be prioritized.**  **Proposal 8: Postpone the discussion on detailed description of use case to next meeting.**  [AC] OK |
| ZTE | General OK to all Proposals proposed by the moderator.  For Proposal 2: Both the wording “processing point of view” and “functional point of view” is fine for us.  For Proposal 3: Since the definition of ML training in the current TR37.817 includes online and offline training, it’s reasonable to change “Model training” to “Model training (offline/online)”. No strong opinion on whether to change “data sources” box in the figure.  For Proposal 5: During the first round discussion, majority of companies think that feedback from action to model training is needed, but no consensus on whether model need this feedback directly, because some companies think the feedback can be achieved from data collection. Hence, we think that the statement can be reworded that “**Feedback from action to model training is needed, and how to reflect it into the AI framework figure is FFS.**” |
| Deutsche Telekom | Ok for all proposals except of Proposal 2:  We agree with Ericsson’s proposal to change “ … from processing point of view” to “from a functional point of view”, as we consider here the functional framework.  Comment to Ericsson’s statement w.r.t. Proposal 3:  As we discuss the functional framework, the term “Data sources” doesn’t make any sense. “Data collection & preparation” relates to the functional task of that block in the figure. We have also still the preference to add “(offline/online)” to “Model training”, otherwise there is a need for 2 separated blocks in the figure. |
| Samsung | In general ok for all proposals except of Proposal 3:  For proposal 3, the definition of “data collection” has been captured in TR, but what “preparation” would do is not clear. Therefore, we prefer “data collection”. If changing “data collection” to “data collection & preparation”, it is better to have a definition or description of “data preparation” in TR.  For proposal 2, fine for both “processing point of view” and “functional point of view”. |
| Intel | Proposal 2: We don’t see a clear difference between “functional pov” and “processing pov”, since each logical node is described with it’s function, slightly prefer “from functional pov”.  Proposal 5: We are fine with current statement in Proposal 5.  Logical node “data collection and preparation” can hold the functionality of the storage and preparation of all types of data which are required for training/re-training. The performance feedback may include all feedbacks and reports as the result of choosing one ML model and taking certain actions according to the inference result. This can also include ML model rewards, reports from other network nodes/UEs based on the taken action, etc. Those data can feed the need for model refining in reinforcement learning or online training by sending data from data collection/preparation node to model training in real time.  If performance feedback is sent to both model training and data collection & preparation, the data is duplicated at two logical nodes, which is not necessary.  Considering that and keep the framework tidy by avoiding duplicate feedbacks to different logical nodes, **we think the feedback to model training is not needed**. |
| Nokia | OK on all proposals proposed by the moderator.  Regarding proposal 2, we realize that functional and processing point of views have become misleading terms causing confusion on the exact meaning. Functionality point of view was meant to be the “training host” for example as in Alternative 1 while processing point of view is the “training” as in Alternative 2. We have already captured in the TR that “The study focuses on AI/ML functionality and corresponding types of inputs/outputs”. Therefore, it is questionable whether proposal 2 provides useful information. One alternative could be that we do not capture this proposal but we capture instead its essence by updating the Figure (to illustrate Alternative 2). |
| InterDigital | For Proposal 2 we think that function pov is better than processing pov but either is ok |
| **Huawei** | **In general, we are fine with those proposals.**  **Some understandings from our side, we also prefer “from functional point of view”; we are also fine with “data source”; for feedback, we could further discuss the feedback from action to model training.** |
| NEC | Proposal 1:  Agree.  Proposal 2:  Our understanding is also that functional framework is described “from functional point of view.” However, because this is not directly reflected in TR maybe we do not need to decide now. This could be discussed later based on concreate TPs to TR.  Proposal 3,4:  Neutral.  Proposal 5:  Agree with ZTE that “majority of companies think that feedback from action to model training is needed, but no consensus on whether model need this feedback directly.”  We think that ZTE wording reflects such view better: “**Feedback from action to model training is needed, and how to reflect it into the AI framework figure is FFS.**”  Proposal 6, 7, 8:  Agree. |
| Lenovo, Motorola Mobility | We are fine with all proposals.  Proposal 3: we agree with Deutsche Telekom. We prefer to use ‘Data collection & preparation’.  Proposal 5: we have different understanding with ZTE. The key point is whether direct feedback from action to training is supported. To us, the answer is no, as explained in the phase I. We insist on the existing proposal 5. |
| CMCC | For proposal 2, similar view as Nokia and NEC, if the terms cause misleading, we perhaps don’t need the proposal and remove it from the agreement, we can just refer to the figure of AI framework in the TP. |
| DOCOMO | Proposal 1. Agree the TR 37.817 v0.1.0  🡺OK.  Proposal 2: AI framework is described from processing point of view.  🡺Among “Functional point of view” and “Processing point of view”, either way of naming is fine.  Alt.2 architecture is preferred since it is simpler and covers all deployment scenarios.  Proposal 3: Change “Data collection” to “Data collection & preparation”, and change “Model training” to “Model training (offline/online)”.  🡺 “Data source” is slightly preferred since the data can be from different source entities.  “Model(online/offline training) is better” since anyway both should be included.  Proposal 4: Feedback from action to data sources is performance feedback.  🡺OK with this terminology. But the content should contains not only "performance" but also other information, such as reports.  Proposal 5: The direct feedback from action to model training is needed.  🡺Slightly prefer the feedback from action to Data Source, instead of model training.  Proposal 6: Postpone the discussion on other open issues proposed by R3-210617.  🡺OK.  Proposal 7: The use cases agreed to start from at RAN3#110 E-meeting could be prioritized.  🡺OK. Should not prevent other use cases, such as optimization in PHY and MAC layers.  Proposal 8: Postpone the discussion on detailed description of use case to next meeting.  🡺OK. |

In the 2nd round discussion, rapporteur would like to discuss the draft TP for TR 37.817 based on above proposals.

**Question 2: Companies are invited to provide views on the draft TP for TR 37.817 (in the draft folder):**

|  |  |  |
| --- | --- | --- |
| **Company** | **Agree or not** | **Reasons/Comments/Suggestions** |
| **ZTE** |  | **Change the Editor’s note: “Whether model training achieves feedback from action directly is FFS.”** |
| **Ericsson** |  | The box “Data Collection and Preparation” implies that:   * There is a single point of data collection, which is not the case; * There is a node/function that can “prepare” inputs for the model, which implies that the node/function has knowledge about the model, when we agreed that the model is up to implementation   We propose to maintain “Data Sources” which implies the following:   * The box represents the set of nodes and functions that may provide input data to the model, which might be a multitude and not one, and that does not necessarily have knowledge about the model implementation   In general we propose to have descriptions for each box, otherwise there are multiple interpretations that can be derived from one name only.  We prefer to maintain the naming “Model Training” and discuss at the next meeting the implications of online and offline training. For example, if both processes are in one box, do we mean that the same node performs both processes? What role does the “Model Deployment” harrow play?  We would prefer to combine Actor and Subject of Action because there seem to be nothing we gain in keeping them separate. RAN3 needs a model that can map to an architecture and from an architecture point of view a function (actor) handles the “subject of action”  We would prefer to remove the “Performance Feedback” harrow and describe the “Data Sources” box as “a function that can provide different types of input data, such as raw measurement, model prediction feedback, inputs required by the training and model functions, prediction of required metrics,…” If this is not possible, we propose to keep the FFS on feedback from “subject of action” to “data sources”. |
| NEC |  | Agree with ZTE proposal.  This wording reflects majority view after the first round of discussion.  Current wording in TP for TR 37.817 “**whether feedback from action to Model training is needed**” does not reflect the majority view correctly.  Regarding combining Actor and Subject of Action: in the first round of discussion there was split 7 by 6 companies. It is better to discuss during the next meeting. |
| Nokia |  | We could like to comment on Ericsson’s following statements:   * Data Collection and Preparation: It is our understanding that Data Collection and Preparation is shown as a single box to illustrate functionality. Then implementation will determine whether you need one or more of this functionality. This block will also need to communicate with the model or have some understanding about the model. Otherwise, how can you prepare the data as needed? * Regarding calling the box “Data Collection and Preparation” versus Data Sources, if we choose Alternative 2 then we would need to also change the naming to reflect processing, namely “Data Collection” (“or Data Collection and Preparation” if this is agreed). In our view, it should by no means be understood that there is a single point of Data Collection. * We also agree that it would be good to describe what each box means for clarity. * The reason why we would like to Keep Actor and Subject of Action separated is that they can reside at different network entities.     **Proposal 4:** One alternative about Proposal 4 is to replace Performance Feedback from Action to Data Collection with Model Performance Feedback (as also indicated by DT and Intel) and replace the arrow from Data Collection to Model Training with “Model Performance Feedback + Training Data”. In this way, Data Collection could be enhanced to receive also ML model rewards and reports from other network nodes based on the actions taken. Also, in our view Performance Feedback is nothing special, just a traditional way of collecting PM and KPIs.  **Proposal 5:** We support that feedback (Model Performance Feedback) from Action to Model Training is needed. However, we think that the Model Performance feedback from Model Inference to Model Training should be removed. At the time of inference, in principle there is no good understanding about the model performance. This is done after the action is taken when the model can be evaluated.  This framework discussion further made us wonder whether we should introduce an arrow “Data Request” from Training to “Data Collection” requesting data for training since currently training entity does not have the opportunity to request data directly from Data Collection. This happens only after an action has been taken. |
| **CMCC** |  | We have some comments on Ericsson’s proposals.   * Data Collection and Preparation: It is just described in a single box, it does not mean there is only a single point of node to collect the data, it could be CU, DU or OAM, etc. From this point of view, data collection and data source has not much difference * Whether to combine Actor and Subject of Action: according to phase I discussion, the views are split, so we could keep the FFS.   ZTE’s proposal to change the editor note, we are fine |
|  |  |  |
|  |  |  |

# 4 Discussion

## 4.1 TR 37.817 v0.1.0

The TR [1] is updated based on the agreements on RAN3#110 E-meeting. It has been checked over the email reflector after RAN3 #110e meeting and resubmitted to RAN3 #111e for agreement.

**Proposal 1. Agree the TR 37.817 v0.1.0**

## 4.2 High-level AI framework

Following are open issues left for AI framework:

*Editor's Note: the details for the framework below is FFS including whether Actor and Subject of action should be in one box or separate, whether feedback from action to Model training host is needed, the name in each box is from functionality or from processing point of view, the feedback from Subject of action to the Data sources is Performance feedback or Model performance feedback and other possible refinement.*

Paper [2][3][4] address on these open issues.

### **4.2.1 AI framework from functionality or from processing point of view**

In R3-210917, two alternatives for illustration of the AI functional framework are discussed:



Figure 1 Alternative 1: AI framework from functionality point of view (currently captured in the TR 37.817)



Figure 2 Alternative 2: AI framework from processing point of view

After some comparison, it is found that the two alternatives do not differ two much. It is proposed that if we cannot reach consensus at this stage, we could keep the current alternative as in the TR and refine it at later stage when we found it is not suitable for use case and solution description.

On the other hand, it is pointed out in R3-210617 that, currently in Figure1, each box represents one processing host to enable AI functionality. However, for one AI functionality in one use case, multiple processing hosts may be used. On the other hand, there is a possibility that more than one AI functionality can be supported by the same one processing host. Therefore, it is preferred for each box to represent one processing action to enable AI functionality including data collection, model training, model inference and action.

Companies are invited to provide views on whether to keep the current alternative as in the TR or choose alternative 2:

**Q1: Do you agree to keep the current alternative as in the TR or choose alternative 2?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Prefer AI framework from processing point of view.  (Alternative 2 is partly OK) | The framework we proposed as follow:    The framework aims to show the entire the ML operation process for RAN intelligence. The framework should be separated from the RAN architecture (including RAN logical node). We prefer the framework comprising four boxes including data collection, model inference, model training and action.  The difference between alternative 2 and framework:  - Performance feedback from action to model training.  - Clarify that model training includes the online training. |
| Deutsche Telekom | We are ok to go with alternative 2 which focuses on functional blocks only (i.e., leaving hosts out of the figure). | With alternative 2 we can address a pure functional AI framework without addressing any deployment aspects. Based on that functional framework, use case specific deployments can be considered in a later phase of the study.  Following proposals for updates of Figure 2 (except of those discussed under following sections):   * Change “Data collection” to “Data collection & preparation” to make clear that this function is not simply collecting data, but also separate it and possibly pre-process it for different purposes. * Change “Model training” to “Model training (offline/online)” to make clear that this function may address both variants.  Note: In a use case specific deployment the function may be split into an offline and an online part which may be placed in different network nodes. |
| China Unicom | We are OK with ZTE’s proposal. | We are fine with the processing model for RAN side. The modifications in ZTE’s proposal are essential for Model training. |
| Nokia | We think both options are OK but | Alternative 2, representing the processing action is more simple since it does not address deployment aspects, e.g., related to where inference or training will be hosted.  We agree with Deutsche Telecom that we need to update Data Collection to “Data Collection & Preparation” to capture not only the process of collecting data, but also the process of preparing/pre-processing the input data for training and inference. |
| vivo | We are fine with Alt2 | Alternative 2 is much focused on AI functional description. We think whether to add “including online training” as proposed by ZTE is just adding additional description for the “Model training”, we are open to consider such addition. |
| Huawei | Fine with Alt 2 | Similar view as mentioned above, each block in alt2 reflects more as functionality, we are also ok with DT’s suggestion. |
| NEC | Both alternatives need some modifications | We think that Actor and Subject of Action should be kept separate.  Do we need multiple Subjects of Action?  We support proposal to have feedback from Subject of Action to Model Training.  How to capture that output from one model could be input to another model? |
| Qualcomm | Slightly prefer Alt 2 | Both alternatives are fine. Using a single box for action is clear. |
| Lenovo and Motorola Mobility | Alt2 as starting point | We agree with DT and Nokia’s understanding. First, we should remove ‘’host’’ since the framework should be functional AI framework. Second, we need to change “data collection” to “data collection & preparation” or “data broker” that is also responsible to prepare data for training/inference as demanded |
| Samsung | Both alternatives are fine. | Alternative 1 is made up of the functionality blocks (not deployment aspects) and relevant interaction between each functionality block. It clearly shows the function & the input/output of each component and the working relationship among all participators, which is beneficial for exploring the network node functionality and interface impact.  Alternative 2 shows the process of AI/ML operation, the functionality of each step and the related input/output of each step, which is beneficial for exploring the lifecycle management of the AI function implementation. |
| CMCC | No very strong view, we can go to alternative 2 | Since this question is related to some of the questions below, for the sake of progress, we could first decide the basic framework, alter 1 or alter 2, further modifications on the two alternatives could be based on the discussion and decision of the following questions. |
| Intel | We are ok with alternative 2 as long as it also allows to allocate each logical functional block to one or more entities or functions. | AI framework in alternative 2 focuses on logical functional blocks of AI-enabled framework. In this case, it is considered that there’s no limitation of supporting functionality of single logical block (e.g. action) to be distributed into different subjects/entities.  Moreover, we agree with DT that we should consider both offline and online training to be supported in AI framework, since there’s no technical limitation preventing offline training to be supported, where study the data collection gaps/interface impact for offline training is also within WI scope. We can explain the meaning of each logical functional block by terminology definition, e.g. training includes both offline and online training. |
| Ericsson | Alternative 2 with modifications | We are fine merging Actor and Subject of Actions into one single entity.  We would like to maintain the Data Sources entity. This is because data may come from different sources rather than been collected in a single entity for data collection.  We have the following further modification proposals:   * Rather than having a harrow for performance feedback, we suggest to describe the Data Sources box in further details, specifying that such data can be measurements from different parts of the network, model predictions, feedback on model prediction accuracy (performance), etc. Having a single harrow terminating in the Data Sources for “performance feedback” would otherwise not be complete. We would need multiple harrows, each for a sub-type of data and from a different part of the system. The proposal of better describing “Data Sources” would also remove the need for a Model Performance Feedback from Model Inference to Model Training * We would like to raise a question about the viability of the model deployment/update harrow. As per our agreements a model details are up to implementation, hence the model is a “black box”. With that it does not seem possible to transfer a model over an open interface, as the content of the model-container will be unknown to the receiver. In light of this, does the “model deployment/update harrow” represent a process internal to a logical node? Do we need this harrow at all, or could we have a single box including model training and model inference, with the understanding that, in practice, these processes may reside in different entities? |
| InterDigital | Alternative 2 + updates | Functional framework should be decoupled from deployment aspects. We agree with updates proposed by ZTE and Deutsche Telekom. |

**Summary:**

**We received 13 company inputs, where 9 companies prefer Alt 2, i.e., AI framework from processing point of view, 3 companies have no strong view, and 1 company thinks both alternatives need some modifications. Therefore, it is proposed to use Alt 2 as starting point since it is the majority view.**

**Furthermore, DT propose to change “Data collection” to “Data collection & preparation” to make clear that this function is not simply collecting data, but also separate it and possibly pre-process it for different purposes. Moreover, change “Model training” to “Model training (offline/online)” to make clear that this function may address both variants. Nokia, Huawei, Lenovo and Intel share the view with DT. Moderator also think the suggestions are acceptable.**

**Moderator’s proposal:**

**Proposal 2: AI framework is described from processing point of view.**

**Proposal 3: Change “Data collection” to “Data collection & preparation”, and change “Model training” to “Model training (offline/online)”.**

### **4.2.2 Whether Actor and Subject of action should be in one box or separate?**

In R3-210917, it is observed that for the purpose of easy identifying the signalling exchange between network nodes, the separation of actor and subject of action seems to be beneficial.

In R3-210785, it is pointed that one box is enough to reflect the function of action. In this framework, the process of the AI should be clearly defined, and the place where model training, model inference, data collection and action (involve actor and subject of action) needs to be discussed case by case.

In R3-210617, it is proposed that to make the Functional Framework simple and straightforward, the separate “Actor” and “Subject of action” boxes can be merged into one block (e.g. Action). The “Action” box does not restrict that only one node or interface is involved for one action, if clear explanation is necessary, one note can be added that “one or more Subjects of Action(s) may act over at least one interface”.

Companies are invited to provide views on whether actor and subject of action should be in one box or separate:

**Q2: Whether actor and subject of action should be in one box or separate?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | The framework aims to show the entire the ML operation process for RAN intelligence. The framework should be separated from the RAN architecture (including RAN logical node).Model inference executes the trained model to get the output based on the inference data. Then, specific policy needs to be generated based on the output. One box is enough to reflect the function of action. |
| Deutsche Telekom | Separate boxes preferred. | With the separation it is more clear that the model interference does not trigger a direct action, but provides information to a function called “actor” which is responsible for triggering actions in different other functions which may be placed in the same node or other nodes of the network. |
| China Unicom | Yes | If considering the functional framework, one box is simple for description. But If considering network deployment, it is uncertain to deploy in one node or separate nodes. |
| Nokia | Separate boxes | Since actor and subject of action may be located in different entities, it is good to be able to capture this in the framework. This separation might also reveal standardization impacts over the interfaces depending on the use case. |
| Vivo | Separate boxes | For more coherent picture of actor and subject, we prefer to have separate boxes. |
| Huawei | One box | From functionality point of view, one box should be enough, whether the action would be taken by two separate entities, this would depend on the concrete use case where whether two entities should be involved or not. |
| NEC | Separate boxes | Actor and subject of action could be located in different RAN nodes or functional entities. Having them in separate boxes is helpful to study standardization impacts on interfaces. |
| Qualcomm | One box | This figure just shows the big function split and workflow. One box for action is clear. This does not prevent from supporting action in multiple network nodes. |
| Lenovo and Motorola Mobility | Yes, one box | To make the Functional Framework simple and straightforward, the separate “Actor” and “Subject of action” boxes can be merged into one block (e.g. Action). The “Action” box does not restrict that only one node or interface is involved for one action, if clear explanation is necessary, one note can be added that “one or more Subjects of Action(s) may act over at least one interface”. |
| Samsung | Separate boxes | The separation helps to investigate standardization impacts over the interfaces depending on the use case. |
| CMCC | Separate boxes slightly preferred. | In some cases, the specific node which is subject to the action may be not the same as the node hosting the actor.  For the purpose of easy identifying the signalling impact between network nodes, the separation of actor and subject of action seems to be beneficial |
| Intel | One box | As commented above, the boxed shown in the figures shows logical node, it is possible that one logical node is distributed on different network nodes. It would be good to have an editor notes to explain that one or multiple subjects can be used as execution units to support processing at one logical functional block. |
| Ericsson | One Box | One box is sufficient to represent the entity that receives the outputs of the model and that uses them according to its policies. Whether using the outputs implies triggering an action towards a different entity, this depends on the use case and it is not central to a study on AI. |
| InterDigital | One box | From a functional framework point of view, one box is sufficient. However, it should be clarified that the functional framework doesn’t imply a constraint on the deployment. A single function may be placed across two different entities, in which case there may be interface impacts. For example, even though the model training function is represented as one box, it does not exclude a distributed training deployment. |

**Summary:**

**We received 13 company inputs, where 7 companies prefer one box, and 6 companies prefer separate boxes. Therefore, no consensus can be achieved at this stage.**

### **4.2.3 Whether feedback from action to data sources is performance feedback or model performance feedback**

Both R3-210917 and R3-210785 think the name “Model performance feedback” is appropriate. Since ML inference is a process of using a trained ML model to make a prediction or guide the decision based on collected inference data and ML model. The output can be feedback to the model training host to verify the performance of the ML model and in turn help the model training host to improve or re-select the ML model.

Companies are invited to provide views on whether feedback from action to data sources is performance feedback or model performance feedback:

**Q3: Whether feedback from action to data sources is performance feedback or model performance feedback?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Feedback from action to data sources is performance feedback.  Feedback from inference to training is model performance feedback. | Feedback from action to data sources is performance feedback.  For the feedback between Model inference and Model training, since Model inference is one component which execute the trained model to get the output based on the inference data, this performance reflects the ML model performance is good or not. If this model performance feedback is not good, ML model needs to be re-selected or re-trained in the Model training component. So feedback between Model training and Model inference is called “Model performance feedback”. Taking AI-based energy saving as an example, the output of the model inference may be the predicted load of the cell.  For the feedback from Action to data collection, this feedback reflects the network performance after the Action. After the policy is adopted , the network performance may be optimized/maintained/degraded. Taking AI-based energy saving as an example, the feedback from Action is the KPI of the network. |
| Deutsche Telekom | Use of “Model performance feedback” is supported. | This term was already proposed in our contribution to the last meeting (see R3-206197). |
| China Unicom | Prefer performance feedback | The data types for performance feedback are different from the data for model performance feedback. It is preferred to have two definitions and performance feedback is the feedback from action to data sources/data collection. |
| Nokia | Performance Feedback | Model performance feedback relates to the accuracy of the model or prediction quality. Feedback from Action to Data Source is related to the impact of the action taken given the model output and is related to performance. |
| vivo | Performance Feedback | Performance feedback to allow assessment of model performance for further training enhancement. |
| Huawei | Performance Feedback | Model performance feedback, our understanding, is to evaluate the trained model accuracy, while the performance feedback is to evaluate the practical effect of the action from the trained model. |
| Qualcomm | Performance feedback | Agree with ZTE |
| Lenovo and Motorola Mobility | Performance Feedback | Agree with China Unicom |
| Samsung | Performance feedback | The feedback from action to data sources is to transfer the network performance after applying AI-generated or AI-assisted policy, such as QoS or other KPI parameters. The network performance data can be used as input to training and/or inference with the aim as 1) feedback the model decision impact for RAN so as to improve the model efficiency; 2) describe the current network status.  For the feedback from inference host to training host is the model evaluation value to evaluate whether model works well (e.g. the accuracy to evaluate a prediction model), which is more proper to be named as model performance feedback.  Thus, we prefer to keep performance feedback to name the feedback from action to data sources. |
| CMCC | Performance feedback | Agree with ZTE |
| Intel | Model performance feedback | To our understanding, model performance feedback includes all feedbacks and reports as a consequence of choosing a ML model and taking certain actions which are inferenced from trained model. It includes both ML model rewards and reports from other network nodes/UEs based on the taken actions. |
| Ericsson | Needs further discussions | It is not correct/exhaustive to have one harrow for performance feedback from Actor to data sources because the actor could provide, as data to be used for AI, also information that are not indication of performance. The actor could for example provide UE measurements as inputs to data sources and measurements could simply indicate radio condition rather than system’s performance. “Performance” is a measure of how well a certain process is doing, while data can be more variegate than that. Hence, we propose not to have a “Performance Feedback” harrow but to work on the description of what are the data types contained in the data sources. The same is valid for the Model Performance Feedback harrow from Model Inference to Model Training |
| InterDigital | Performance Feedback | We think that the ‘Action’ function may not have complete knowledge about the AI model. So, the feedback from ‘Action’ to ‘Data Source’ is related to the performance feedback. |

**Summary:**

**We received 12 company inputs, where 9 companies prefer to use performance feedback, 2 companies support model performance feedback, 1 company suggests further discussion. Therefore, it is proposed to use performance feedback since it is the majority view.**

**Moderator’s proposal:**

**Proposal 4: Feedback from action to data sources is performance feedback.**

### **4.2.4 Whether feedback from action to model training host is needed**

Some company deems that performance feedback from action to data source is enough [2], but others find that the feedback from action to model training is needed for re-training or reinforcement learning [3].

Companies are invited to provide views on whether feedback from action to model training host is needed:

**Q4: Whether feedback from action to model training host is needed?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | Model training is one component that train the ML model offline or online according to different use case, so re-training or updating model is also one part of model training. Model training may be triggered to retrain the ML model when the model performance feedback is not good. Similarly, if the network performance feedback from Action is not good, Model training should also be triggered to retrain the ML model or update the ML model.  On the other hand, reinforcement learning, as one of the ML techniques (online training), needs to interact with the network environment during training phase. During training phase, the agent in reinforcement learning needs to get the reward from the environment through performance feedback from action. If the framework should support the reinforcement learning, the performance feedback to the model training is essential. |
| Deutsche Telekom | Yes, but … | Taking alternative 2 as baseline for functional framework where “Model training” should cover both online and offline training, the direct feedback loop is only required for online training (optional link). This can be clarified by using e.g. a dashed line in the figure and additional text for explanation.  Re-training/reinforcement learning aspects have to be considered in use case specific deployment scenarios where both training schemes may be in different hosts/nodes and therefore the data flows may be also split. |
| China Unicom | Yes | We agree to implement online and offline training, and the feedback from action to model training host is necessary |
| Nokia | Yes | Feedback from Action to Model Training can be useful for online training and reinforcement learning. |
| vivo | Yes | Agree with China Unicom. This is necessary for online training and ML enhancement. |
| Huawei | Yes | Similar view as Nokia. |
| NEC | Yes | This is needed for online/reinforcement training. |
| Qualcomm | Yes | It is useful for online training and reinforcement training. |
| Lenovo and Motorola Mobility | No, we think the feedback should be from ‘action’ function to ‘data collection and preparation’ function. | We don’t believe it’s practical to provide all the available data all the time to the AI training function. **Even if we say Action can deliver data (e.g. performance feedback) for AI training, we are assuming part of Action has the “data collection and preparation” function.** Thus, we prefer to let “data collection and preparation” function play the role of data collection from all functions and deliver the demanded data for AI training/inference. |
| Samsung | NO | There are multiple factors to affect the network performance, such as channel condition, generated policy, equipment issues etc. If the network performance is bad, the analysis is required to be done to find out the reason firstly. When the problem is coming from AI model based on analysis, the model needs to trigger the retraining procedure, otherwise, the model-retraining should not be triggered. So the retraining procedure cannot be triggered directly by the network performance.  For the reinforcement learning, the performance data can be one of the training inputs. For the current framework, the performance data can be obtained by training host from data sources as training data, which is more reasonable for the role of performance data for reinforcement learning.  Thus, we prefer the framework without feedback from action to model training host. |
| CMCC | Yes, partly | Feedback from Action to Model Training is useful for online training and reinforcement learning. Further question is whether it get the feedback directly from action or from the data collection |
| Intel | No | Data collection is a logical node responsible for all types of data collection, which can be treated as a data source pool. “Data collection” logical node can transfer “model performance feedback” to “model training” node based on requirement. “model training” can further re-train ML model if needed. There’s no need to duplicate model performance feedback to “model training”. It would be clear from framework point of view that we keep the all data stored and reported to data collection logical node.  Besides, there’s no limitation for reinforcement learning or online training to be adopted in such AI framework if model performance data only reported back to data collection. Data can be transferred from data collection to model training in real time to support online training and reinforcement learning. |
| Ericsson | No | The “Data Sources” box represents all entities that can provide data relevant to the AI process. One of such entities could be the Actor. Therefore, the current diagram is already including the option for performance feedback to be transferred directly to the model inference function. If we draw a specific harrow from actor to model inference, then we seem to say that “data sources” is a separate new entity, not including the Actor, which is responsible for collecting all possible data needed for AI, which is not so far agreed. |
| InterDigital | Yes | We think that the framework should be generic enough to accommodate different AI approaches. We think feedback from action to model training is needed to enable online (re)training and reinforcement learning. |

**Summary:**

**We received 13 company inputs, where 9 companies think that feedback from action to model training host is needed, and 4 companies don’t agree.**

**Moderator’s proposal:**

**Proposal 5: The** **direct** **feedback from action to model training is** **FFS.**

### **4.2.5 Other open issues for AI framework**

In R3-210617, some other open issues are discussed and following proposals are proposed:

**Proposal 3: RAN3 supports the case that one ML model demands input from other ML models.**

**Proposal 4: For the sake of discussion, RAN3 further distinguishes the decision-oriented ML model and the prediction-oriented ML model when it comes to ML model feedback provision and ML model retraining/update.**

**Proposal 5: The data collection shall provide the training data according to the demand of model training regarding what/when/how to provide. It is regarded as model performance feedback when the training data provision is triggered due to bad model performance.**

**Proposal 6: The data collection also collects the output from the model inference.**

Since above issues have not been touched in last RAN3 meeting, companies are invited to provide views on whether to discuss these open issues in this offline discussion:

**Q5: Whether to discuss above open issues proposed in R3-210617 in this offline discussion?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE |  | P3, P4 may discuss case by case later..  For P5, we think there is no need to add the data provision policy from model training to data collection into the current AI framework. The framework aims to show the entire the ML operation process for RAN intelligence rather than signaling design  For P6, not needed, model training can get the prediction result through the model performance feedback from model inference to model training. |
| Deutsche Telekom | No | Due to 0 TUs in this meeting we should focus on items in Sec. 3.2.1 – 3.2.4 only. |
| Nokia | No | We agree with Deutsche Telecom. |
| vvio | No |  |
| Huawei | Pending on the available TU | Given there is TU available, we could discuss, but the proposals themselves are not very clear, e.g. for P3, is it referring to federal learning? for P4, I suppose it is related Q6? |
| Qualcomm | No | Agree with DT |
| Lenovo and Motorola Mobility |  | Postpone |
| Samsung |  | Same view with DT. Prefer to have the discussion at next meeting with allocated TU. |
| CMCC |  | These details can be discussed in future meetings |
| Intel | No | We agree with DT. |
| Ericsson | No | Wise comment from DT |
| InterDigital | No | Ditto, agree with Samsung, Intel, and E///. |

**Summary:**

**It is observed that all companies that answered the question agree to postpone the discussion due to 0 TUs in this meeting.**

**Moderator’s proposal:**

**Proposal 6: Postpone the discussion on other open issues proposed by R3-210617.**

## 4.3 Use case

### **4.3.1 Use case priority**

In R3-210073, it is observed that most interested use cases are: energy saving, traffic/load prediction, trajectory prediction, traffic steering and load balancing. Moreover, some of them can be used as a tool box by other use cases, e.g., load prediction, UE trajectory prediction. Therefore, it is proposed to study the below use cases in R17 with high priority:

1. Tool box use cases:

- traffic/load prediction

- trajectory prediction

1. System level use cases:

- energy saving

Companies are invited to provide views on whether to study above use cases in R17 with high priority:

**Q6: Whether to study above use cases in R17 with high priority?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | The most interested use cases are: energy saving, traffic/load prediction, trajectory prediction, traffic steering and load balancing.  Some of them can be regarded as system level use case. Take energy saving as an example, model training can be deployed outside RAN, e.g., OAM system, the whole system solution including three main functions, AI based scenario classification, AI based Load prediction, and AI based ES performance analysis.  While some of them can be used as a tool box by other use cases, e.g., load prediction, UE trajectory prediction. For example, load prediction can contribute to use cases like energy saving, load balancing, mobility optimization. UE trajectory prediction can also contribute to use cases like mobility optimization, energy saving. For those tool box use cases which can be achieved purely on the RAN side for RAN purpose. |
| Deutsche Telekom | Yes, but … | We are in principle ok to start with those use cases, but nevertheless, this should not prevent other use cases to be considered within the SI phase (contribution-driven). |
| China Unicom | Yes | The above three use cases are good start for R17, but we should not prevent other use case in the future discussion. |
| Nokia | Not quite | Energy saving is one of the approved use cases to start with and we support it. However, traffic/load prediction and trajectory prediction are not really use cases in our view but are, rather, enablers that could be used by other use cases.  We support to start by the approved use cases of energy saving, load balancing, traffic steering/mobility optimization. To allow some diversity in the cases we consider, we also support use cases for optimization of lower layer parameters. |
| vivo | Yes | We agree to support all above use cases with high priority. From our understanding these are basic scenarios subject to AI. |
| Huawei | Yes for (2) | Similar view as Nokia, for energy saving for sure yes. For (1), we think they are just intermediate steps which would be applied to different use cases, such energy saving, load balancing, mobility optimization, etc. |
| NEC | Yes for energy saving | We support to start from use cases selected in 110e but allow to consider other use cases case by case. |
| Qualcomm | Yes | Energy saving has been agreed.  Tool box use cases can provide input and support for other use cases. So, 1) should have higher value than the system level use cases. |
| Lenovo and Motorola Mobility | Yes | We agree to study the use cases such as energy saving, load balancing and traffic steering/mobility optimization in R17 with high priority. |
| Samsung | Yes for (2) | Prefer the agreed use case in 110e meeting to be the start point. And discuss the other use case at next meeting with allocated TU. |
| CMCC |  | The use cases we agreed to study at last RAN3 meeting could be priorotized. |
| Intel | Maybe | It is agreed in last meeting that we continue study potential new use cases and other use cases are not precluded. We also achieved the agreement to focus on “energy saving, load balancing, traffic steering/mobility optimization” as a starting point. Hence, we don’t think this proposal is needed. |
| Ericsson |  | We back agreements at RAN3-110e to work on energy saving, load balancing, traffic steering/mobility optimization. As said, this is not excluding other use cases. We see no actual benefit in labelling use cases as “toolbox” or “system level” |
| InterDigital |  | We think that traffic/load prediction and trajectory prediction are building blocks which can be applied for more than one use case. One approach is to identify building blocks for each use case and prioritize the building blocks which are common for the most use cases.  As a starting point, we should consider the list of use cases agreed to be studied during the last meeting. |

**Summary:**

**It is observed that all companies answered the question agree that energy saving should be study with high priority since it has been agreed on last meeting. As for traffic/load prediction and trajectory prediction, some companies think they are not really use cases.**

**Furthermore, in last RAN3 meeting, we have agreed to start from energy saving, load balancing, traffic steering/mobility optimization and other use cases, e.g., optimization of physical layer parameters, are not precluded. Therefore, we could start from the use cases agreed in RAN3#110 meeting.**

**Moderator’s proposal:**

**Proposal 7: The use cases agreed to start from at RAN3#110 E-meeting could be prioritized.**

### **4.3.2 Use case description**

R3-210918, AI based Energy Saving is used as the starting point and intend to work out an example on how to describe the use case and solutions. Companies are invited to provide views on the description if any:

**Q7: Do you agree to use Energy Saving as the starting point and what’s your comments about the description in R3-210918 if any?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | The detail use case description can be further discussed in the next meeting. | As we analyzed above, there are three main functions for AI based energy saving (AI based scenario classification, AI based Load prediction, and AI based ES performance analysis), for each function, how to deploy the AI framework into the current system architecture, and the detail inputs and outputs are different, which needs to be further discussed. |
| Deutsche Telekom | No | Due to 0 TUs in this meeting we should focus on items in Sec. 3.2.1 – 3.2.4 for the functional framework only. Going into details of a use case description does not make sense without a stable definition of a functional framework that can be used as a basis. |
| China Unicom | The detail description can be further discussed. | AI based energy saving is a good use case, but the detail of the description of that could be discussed in next meeting. Before we discuss the use case, the framework needs to be defined at the first stage. |
| Nokia | No | We should first focus on the ML framework and then go into details of specific use cases. We cannot study use cases before having defined a solid functional framework. |
| vivo |  | Energy saving should one of the use cases to start with, but not the only one. We think description of this use case should be further considered. |
| Huawei | Yes | Anyway we should focus on some use cases, now we saw more than ten use cases, if we have no focus, it is difficult to converge I am afraid. And, energy saving is one of the most useful features contributing OPEX reduction which is so important to operators. |
| NEC |  | Detailed use cases could be discussed in the next meeting. |
| Qualcomm |  | Agree with ZTE. The description can be discussed in next meeting based on contributions. |
| Lenovo and Motorola Mobility | The detail description can be further discussed. | Energy saving is one use case that to be studied in R17, but the details can be postponed. |
| Samsung |  | Detailed use case can be discussed in the next meeting. |
| CMCC |  | We can focus only on AI framework at this meeting. R3-210918 could be reference for use case and solution description in the subsequent meetings |
| Intel | No | We don’t have sufficient time in this meeting to discuss or agree on the baseline of any use case. |
| Ericsson |  | Let’s simply agree to the use case prioritisation established in RAN3-110e and leave the details for the next meeting. |
| InterDigital | Use case specific aspects can be discussed next meeting | Due to limited time unit allocation this meeting, we prefer to make progress on the functional framework. A stable functional framework can then be used as basis for use case specific discussion. |

**Summary:**

**It is observed that all companies agree to postpone the discussion on detailed description of use case to next meeting.**

**Moderator’s proposal:**

**Proposal 8: Postpone the discussion on detailed description of use case to next meeting.**

# 5 Conclusion, Recommendations

To be edited, if needed**.**

# 6 Reference

1. R3-210935, TR 37.817 v0.1.0 (CMCC)
2. R3-210917, Further discussion on high-level AI framework (CMCC)
3. R3-210785, High-level framework and definition for AI RAN (ZTE Corporation)
4. R3-210617, Open issues of framework for AI (Lenovo, Motorola Mobility)
5. R3-210918, AI based Energy Saving (CMCC)
6. R3-210073, Use cases for AI study (ZTE Corporation)