[RAN94e_R18Prep-08] AI/ML for Air Interface - Version 0.0.6
RAN

3GPP TSG RAN Meeting #94-e **RP-212668**

Electronic Meeting, December 6 - 17, 2021

Agenda Item: 8A.1

Source: Moderator (Qualcomm)

Title: Moderator's summary for discussion [RAN94e-R18Prep-08] AI/ML for Air Interface

Document for: Report

1 Introduction

[1] identified the following topic for further discussion after the Rel-18 RAN workshop:

- **12. AI (Artificial Intelligence)/ML (Machine Learning)**, with the following example areas:
  - Air interface (e.g., Use cases to focus, KPIs and Evaluation methodology, network and UE involvement, etc.)
  - NG-RAN

Subsequent NWM discussions took place during the period August 30th – September 3rd and were captured in [2].

While no formal agreements were made, discussions in this thread pick up were relevant discussions on AI/ML for Air Interface left off. As a result, the moderator suggests focusing on the following areas:

- Candidate project properties
- Assumptions incl. nomenclature, evaluation methodology
- Use Cases for AI/ML for Air Interface project
- KPIs
- 3GPP framework for AI/ML for Air Interface
- Other

The end goal of this discussion is to generate a candidate SI for Rel-18 as indicated in [3].

2 Initial Round

2.1 Candidate Project Properties

Based on the discussions in [2], the project on AI/ML for Air Interface would have the following properties:

Table 1:

Title	SI/WI	Primary WG	Secondary WGs	Notes
AI/ML for Air-Interface	SI	RAN1	RAN2, RAN4	SI for entire Rel-18 duration. No foreseen impact to/from SA.

Company inputs on candidate project properties are welcome in the following review form:

Feedback Form 1: Candidate project properties

1 – vivo Mobile Communication Co. We are fine in principle except the note that ”no foreseen impact to/from SA”, this may depend on potential discussion in both SA and RAN. We don’t think it is appropriate to directly state like this before situation is more clear.
2 – NTT DOCOMO INC. We are fine with candidate project properties shown in Table 1.
3 – LG Electronics Inc. Support the proposal
4 – CAICT We are generally fine with the proposal. The note seems unnecessary before we have clearly definition of each use cases.
5 – Spreadtrum Communications Regarding the Note part, we prefer to delete ‘No foreseen impact to/from SA’. It perhaps depends on the subsequent thoroughly discussion during SI phase. It is too early to set the bar. For other parts, we support Moderator’s views.
6 – ZTE Corporation Agree to have an SI spanning the entire Rel-18 duration. Agree to have RAN1 as primary WG and RAN2/4 as secondary WGs.

7 – AT&T

We agree with the moderator's proposal.

8 – Guangdong OPPO Mobile Telecom.

In general we are fine. But we think it is premature to say "no foreseen impact to/from SA". As we pointed in August email discussion, the exchange of AI/ML model as well as some relevant training/inference data may demand quite high data throughput which should be also taken into account

9 – SHARP Corporation

We agree with the proposed project.

10 – FGI

We agree with the proposal except for "No foreseen impact to/from SA". It seems too early to conclude that there is no impact to/from SA.

11 – CATT

Support the proposal.

Regarding note, we think'No foreseen impact to/from SA' accurately reflects current status, as the impact cannot be clearly stated by now. Depending on further discussion, if some impacts are identified, the SID can be updated to reflect the impact.

12 – Sony Group Corporation

We are fine with candidate project properties.

13 – HUAWEI TECHNOLOGIES Co. Ltd.

Agree that AI/ML for Air-Interface should start from SI with RAN1 as leading group, spanning entire Rel-18 duration. It is not clear yet whether to list RAN4 as the secondary WGs, some further discussion/clarification on what work expected for RAN4 may be needed. Based on the discussion in the previous meetings, it seems some companies think RAN4 need to define the performance requirement, but does this belong to a WI phase?

14 – Rakuten Mobile

We are generally fine for the project property except the Notes. It might be too early for ensuring no foreseen impact to/from SA.

15 – DENSO CORPORATION

Agree with the proposed properties, albeit not sure if SA impact is not foreseen during the entire SI phase.

16 – Samsung Electronics Co.

We are ok with table 1.

17 – MediaTek Inc.

We'd like to clarify RAN2 scope, based on the listed use cases below, only R1 functions are included. RRM mobility was mentioned in previous round, but not listed now; if this is the case, we'd suggest to remove R2 from the secondary WG, so no R2 TU is required.

18 – BT plc

Duration and WG composition look reasonable to use

19 – Lenovo (Beijing) Ltd

Agree to have such SI on AI/ML interface led by RAN1 as primary WG and RAN2/RAN4 as secondary WGs.

Furthermore, following issues need be considered when formulating the objectives/scope:

- Entire Rel-18 duration for SI
- Involvement of RAN4 to determine the requirements and new testing methods for the new identified KPIs evaluation could be needed in following normative work.
- An explicit note related to SA's involvement is not needed.

20 – NEC Corporation

we support this proposal.

21 – Fraunhofer IIS

Support the proposal.

22 – Fujitsu Limited

We support this proposal in general. It seems no need to note "No foreseen impact to/from SA" here as pointed by other companies.

23 – ROBERT BOSCH GmbH

Agree with moderators proposal.

24 – Keysight Technologies UK Ltd

We agree in principle with the moderator's proposal. Although we think that the "No foreseen impact to/from SA" note should be removed since we should not make that assumption at this point.

25 – Nokia Corporation

Moderator's proposal is ok for us assuming. We see it important that although RAN1 is primary WG, RAN2 and RAN4 need to start their studies early to ensure that RAN1 studies can take RAN2's and RAN4's findings into account.

26 – VODAFONE Group Plc

OK with table 1 except we might need SA to assist in e.g. provision of some anonymised UE identification, so please remove the part on SA.

27 – China Mobile Com. Corporation We support the proposal in principle. We also think the note“No foreseen impact to/from SA”should be removed, since AI/ML model exchange may be considered, which is relevant to SA.
28 – Deutsche Telekom AG We are generally fine with the proposal except of the statement ”No foreseen impact to/from SA.” This should be deleted as from AI/ML model lifecycle management perspective, it is unclear how initial model deployment and initial training, testing and validation will be handled. The relation to OAM should be clarified in the SI phase.
29 – Intel Technology India Pvt Ltd we support this proposal
30 – Apple Hungary Kft. Agree with the proposal
31 – Ericsson LM We support the proposal.
32 – Futurewei • We agree on setting up a RAN1-led Rel-18 SI for AI/ML for Air-Interface. • We suggest removing the 2nd item on the notes as it’s too early to assume “no foreseen impact to/from SA” before use cases being agreed for the SI yet.
33 – Qualcomm Incorporated **We support the proposal.**
34 – InterDigital We agree with the moderator proposal.

2.2 General: Nomenclature

AI/ML procedures have two well differentiated phases:

- Training
- Inference

For either of the two phases above, the following collaboration frameworks were identified in [2]:

0a) No collaboration framework: AI/ML algorithms purely implementation based and not requiring air-interface changes.

0b) No collaboration framework with modified Air-Interface catering to efficient implementation-based AI/ML algorithms.

1) Inter-node assistance to improve the respective nodes' AI/ML algorithms. This would apply to UEs getting assistance from gNBs (for training, adaptation, etc.) and vice-versa. This level does not require model exchange between network nodes.

2) Joint ML operation between UEs and gNBs. This level requires AI/ML model instruction or exchange between network nodes.

While clearly Cat 0a does not require air-interface changes, it may provide an insightful baseline to assess the (performance/complexity) value of Cat 0b, 1, 2 based solutions.

These categorizations are applicable to training and inference phases independently.

Company comments on nomenclature are welcome in the following review form:

Feedback Form 2: Nomenclature

1 – vivo Mobile Communication Co.

We are fine with the basic principles for above categorization, i.e., categorization based on level of system impact and independent categorization for interference and training.

One more clarification from our side: what is the key difference between 0b) and 1)? The following two aspects need to be clarified.

1. One may interpret the modified air interface in 0b) can also be used to provide assistance information described in category 1). Is it possible to make it more clear for the modified air interface and assistance information in above category description?

2. When the exchanged information is not directly related to model information but related to AI/ML input information that may facilitate AI/ML operation, which category does it belong to? For example, some beam pattern information exchanged for the purpose of beam management in inference stage, or training data provided for training stage?

2 – NTT DOCOMO INC.

The difference between 0b) and 1) ~~is~~ unclear to us. Especially, the definition of assistant data in 1) and the difference between "modification of air-interface" in 0b) and "support of assistance" in 1) are not clear. Although we agree in principle that it would be helpful to have these categorizations, at least further clarification especially for 0b) and/or 1) seems necessary.

3 – LG Electronics Inc.

Although this categorization is fine in general, we wonder whether this categorization is deemed necessary for defining SI objective. Is the intention to limit the SI scope to a particular category? If we proceed SI for entire Rel-18 duration, the category could be selected based on the results of study, i.e. at the end of Rel-18, and the selection could be use case specific. Besides above question/comment, we'd like to clarify Cat1 a bit further. Does this cat mean to signal data set between nodes or to signal UE/gNB capability related to AI/ML algorithm?

4 – CAICT

The collaboration frameworks for training and inference reflect the AI/ML related data and model exchange. The more detail discussions could be included in study phase.

5 – ZTE Corporation

We are generally okay with this nomenclature and categorization. Suggest to clarify Cat 1 as follows.

1) Inter-node assistance to improve the respective nodes' AI/ML algorithms. This would apply to UEs getting assistance from gNBs (for training, adaptation, *inference* etc.) and vice-versa. This level does not require model exchange between network nodes.

6 – Guangdong OPPO Mobile Telecom.

We think there is something different between category 0b) and 1). In 0) the Uu interface is enhanced from the help of implementation based AI/ML, but not vice versa. While in 1) the assistant information over air is to help improve AI/ML algorithm, but without exchange of AI/ML model directly. But we intend to think Uu interface could be also enhanced due to same reason in 0). The same issue is foreseen for 2). Here are our suggestion on 1) and 2):

1) Inter-node assistance to improve the respective nodes' AI/ML algorithms. This would apply to UEs getting assistance from gNBs (for training, adaptation, etc.) and vice-versa. This level does not require model exchange between network nodes. Air-interface could be improved catering to efficient AI/ML algorithms.

2) Joint ML operation between UEs and gNBs. This level requires AI/ML model instruction or exchange between network nodes. Air-interface could be improved catering to efficient AI/ML algorithms.

7 – SHARP Corporation

We agree with moderator's proposal in general. We believe that such categorization is necessary for both Rel-18 SI and studies in future releases, if the group is also considering long-term plans. Different categorizations may lead to very different specification work load. For Rel-18 SI which is the beginning phase of applying AI/ML for air interface, it is beneficial to focus on categorizations with less specification impact. Other categorizations could be studied step by step during the future releases.

For the details, although we understand the intention of distinguishing Cat 0b and Cat 1 (and we agree to distinguish these two), we share the same view that further clarification is needed.

8 – FGI

The differentiation between the collaboration frameworks 0b) and 1) is not clear, especially "with modified Air-Interface". It is confusing that if there is no collaboration, then why the air-interface is involved and modified? If the air-interface is involved and modified, then it seems to be as the same as the description in the collaboration framework 1).
Regarding the term "inter-node" in the collaboration framework 1), it is confusing whether the interaction is between UE and gNB or the interaction involves gNBs. The main reason is that from RAN2's perspectives, "inter-node" RRC message is involved between gNBs.

9 – CATT

We would also like to get a better understanding on the difference between 0b and 1.

10 – Sony Group Corporation

We are generally fine with the categorization, but the categorization should be studied within the study item rather than being some input to the study item.

We think that the meaning of Cat 0b needs clarification. We understood the meaning of no collaboration as "signaling is not involved in training and inference". We understand that 0b could entail changes to the air interface, such as new reference signals.

We understand that the collaboration frameworks would be considered separately for (1) the training phase and (2) the inference phase.

11 – HUAWEI TECHNOLOGIES Co. Ltd.

We doubt the necessity to discuss categorization from defining the scope of SI perspective. Clearly we will not be able to categorize any use case into one of the above 0-2 prior to the knowledge of the potential benefits it may provide, which would depend on detailed discussion/evaluation. In addition, it is difficult to justify whether the above categorization is appropriate or not without thorough discussion/evaluation for each specific use case, by considering the procedures of model training and model inference. In a word, this kind of discussion should belongs to part of the study, not for defining the objective of the study item.

12 – Rakuten Mobile

The clarification is needed for Cat1 and Cat2 whether it is required air-interface changes or not, explicitly.

13 – DENSO CORPORATION

The proposed framework is a good start point for the evaluation in the SI phase. On the other hand, it is not clear if it has to be agreed before starting the SI. Anyway, as already commented, the a clear-cut distinction between 0b and 1 would be helpful.

14 – Samsung Electronics Co.

The categorization looks okay as it covers all the possible cases. Since AI is a new topic in 3GPP, at this stage it may be good to study all the frameworks. For instance, solutions falling in 0a) such as channel estimation, MIMO receiver algorithm, etc., can affect system evaluation assumptions and also may provide insights to solutions falling in other categories at least for the sake of comparison. On the other hand, it is also natural to expect that some preference will be given to the categories having a standard impact, i.e., 1) and 2). Nonetheless, it would be better not to preclude 0a) and 0b), which would help 3GPP to gain a holistic understanding of how AI/ML could interact with NR as a whole.

We have some clarifications that are needed on Cat 1), and Cat 2:

For Cat 1, it is characterized as no model exchange compared with Cat 2, so even data-channel carried model exchange is still precluded for Cat 1? If so, better to clarify.

The model instruction is allowed for Cat 2: For model instruction, is it that possible to pre-configure several AI neural networks at one network node, and the corresponding instruction is followed to use one of the pre-configured AI neural networks?

Model exchange for Cat 2 potentially includes model structure, or model weights, or both structure and weights. Furthermore, the exchange can be used for inference and training.

15 – MediaTek Inc.

For 1) and 2), except signaling between NW/UE and NW/NW, was there any expectation on Air-Interface modification, i.e. similar to 0b)?

16 – BT plc

Nomenclature looks reasonable to us, except for the ”air interface changes” element that seems unclear at the moment - almost certainly we will need some change to go beyond implementation-specific solutions. We suggest to separate collaboration level and air interface change level.

For example, rename the collab. frameworks as

0a) -> 0)

0b) -> remove category

1) -> 1)

2) -> 2)

And define indicative levels of air interface changes

A - no air interface change

B - additional information elements, no other specification change

C - specification change, e.g. new signals

D - …

17 – Lenovo (Beijing) Ltd

It is important to clearly define the collaboration levels of the AI/ML-related operations for any selected use case and AI model/algorithm. Because the model may be transferred or configurable, we prefer to use ‘**Model generation**’ instead of ‘training’ for the used model.

Thus, there are following options for the potential air-interface changes:

- Model generation
  - Initiation/configuration: e.g., to indicate and/or configure which or whether AI model is used
  - Training: e.g., to assist training with dedicated data sets transfer
  - Model transfer: e.g., transfer of all or partial trained model
    (Note: the last above two options both may need additional signals and procedures for the model generation.)
- Inference
  - To infer results and/or better facilitate PHY procedures

Thus, we summarize the categories in a table below with new number-list for each category.

- **Cat.0 (baseline)**: No collaboration, i.e., using existing NR framework as baseline for performance comparison.
  - Generation: Initiation/Configuration: No
  - Generation: Training/Transfer: No
  - Inference: No

- **Cat.1**: No collaboration, but to enhanced PHY procedures and/or signaling assuming AI/ML used in implementation.
  - Generation: Initiation/Configuration: Yes
  - Generation: Training/Transfer: No
  - Inference: No
- **Cat.2**: Inter-node collaboration to enhance PHY procedures and/or signaling to better facilitate inference operation
  - Generation: Initiation/Configuration: Yes
  - Generation: Training/Transfer: No
  - Inference: Yes
- **Cat.3**: Inter-node collaboration to enhance PHY procedures and/or signaling to facilitate model transfer/training and inference
  - Generation: Initiation/Configuration: Yes
  - Generation: Training/Transfer: Yes
  - Inference: Yes

18 – NEC Corporation

We also think 0b) should be clarified.

For 1) we want to clarify if "inter-node" means UE between gNB only, or also include gNB between gNB, UE between UE?

For Cat 2) we suggest to differentiate the following collaboration levels.

2a) Split AI/ML operation. AI/ML model is split into multiple parts and both gNB and UE are involved in training the AI/ML model. For example, auto-encoder-like CSI feedback compression where UE is the encoder and gNB is the decoder.

2b) AI/ML mode distribution. The level means no joint training but exchange the training AI/ML model. For example, UE only downloads the AI/ML model from gNB to predict future CSI.

19 – Fujitsu Limited

Similar view as LG, we wonder whether it is necessary to clarify the categorization in discussions during setting up SI stage? If an SI spanning entire Rel-18 is assumed, this issue can be discussed later under the scope per use case. For each use case, the categorization might need to be defined/refined respectively.

20 – Keysight Technologies UK Ltd

In our opinion, the difference between categories 0b and 1 is not clear. From our point of view, categories 0b and 1 could be potentially merged.

This categorization should also be independent per use case, as some categories may not apply for some use cases.

If, as stated in section 2.3, the evaluated AI/ML schemes should be based on offline training, then only category 0a would apply in the training phase.

We would like to suggest extending the AI/ML procedures nomenclature to cover:

1. Training – learning process & fit model models
2. Validation – tune AI/ML models
3. Testing – assess performance of AI/ML models
4. Inference

We would also like to suggest explicitly defining that training, validation and testing phases can be done with ”synthetic data” and/or ”field data”.

21 – Nokia Corporation

Moderator’s proposal for the categorization is ok, at least as starting before conducting detailed studies, We have similar views as Vivo and NTT Docomo that difference between 0b) and 1) is not clear. It would be good to clarify the difference of 0b) from 1) e.g. as follows:

0b) No collaboration framework with modified Air-Interface catering to efficient implementation-based AI/ML algorithms. The modified Air-Interface may not need any inter-node assistance to improve AI/ML algorithms.

We see that it is important that for different collaboration frameworks RAN4 carefully studies how to ensure predictable UE behaviour with AI/ML and UE to define UE requirements and test cases with different collaboration frameworks. These findings should also be taken into account in studies of different collaboration frameworks and when assessing different collaboration frameworks.

22 – ROBERT BOSCH GmbH

1) We share a similar view as LG. We believe that a categorization itself might be very helpful to give some structure to use case as AI/ML is a fairly novel topic. However, we wonder if the use cases discussed during the study will reveal dependencies not covered by this rather coarse categorization.

23 – China Mobile Com. Corporation

The difference between 0b) and 1) should be clarified. In our mind, both 0b) and 1) need Air-Interface change. We also wonder that whether the categorization is needed at this stage, since some categories may only apply for some use cases, we think the categorization can be discussed together with use case.

24 – Xiaomi Communications

We think there is no strong need to discuss the categorization at current stage. And we are also a little puzzled about what role the categorization plays. Will it be used to assess the impact on the air interface?

We share similar with with China Mobile, the categorization can be discussed with the use case.

25 – Deutsche Telekom AG

We are in general fine with the categories defined, but the terms used in the description needs further explanations. E.g. what is meant by ”collaboration”? What are the players here (e.g. UE and gNB, several gNBs)? And more criteria like those proposed by Lenovo can be integrated as sub-items to specify the meaning.

26 – Intel Technology India Pvt Ltd

Difference between 0b) and 1) is not clear. How should we interpret level 2) model exchange in conjunction with "applicable to training and inference phases independently" ? No problem with the categorization but we think this can be left out of the SI objectives

27 – Apple Hungary Kft.

Need clarification between 0b and 1. It is not clear what exactly the difference between "modified Air-Interface catering to efficient implementation" versus "assistance to improve the respective nodes' AI/ML algorithms". For example, RS overhead reduction due to AI/ML based channel estimator, is 0b or 1.

28 – Ericsson LM

We understand that the collaborative framework is built around implementation-based (proprietary) AI/ML algorithms, following the higher-level principle agreed to in RAN. Moreover, any potential specification activities will be focused on interfaces and signaling.

The categories appear to be a good starting point for discussions; however, the need for them in the SID is not clear. The categories are not completely defined, for example,

- How "development domain" (i.e., outside of the RAN / IODT) based training fits into the categories is ambiguous for some use cases.
- In category 1) inter-node is confusing since it can be interpreted as e.g. gNB to gNB, which we believe is not in the scope.
- It appears that category 1 applies only to training (updating / adapting algorithms).

Even if this discussion is useful and needs to take place, we do not see the need to spend time on formalizing the categories now as this discussion will be carried out in the SI.

29 – Futurewei

• At pre-SI stage, we suggest providing a general categorization based on collaboration but no need to focus on the detailed definitions as the details may evolve during the SI phase when we study the identified use cases.

• In general, we are ok with the above identified collaboration categories of 0a) and 2). However, more clarification is needed for 0b) and 1).

30 – Qualcomm Incorporated

To avoid confusion, we think it is better to discuss inference and training separately. For example, assistance data for inference and assistance training data for training are very different and should not be put into one category.

Therefore, we prefer having the above categorization based on inference only.

For inference:

0) No collaboration framework: AI/ML algorithms purely implementation based and not requiring air-interface changes.

0b) No collaboration framework with modified Air-Interface catering to efficient implementation-based AI/ML algorithms. Examples include a new RS pattern or a new beam pattern.

1) Inter-node assistance to improve the respective nodes' AI/ML algorithms: Signaling, assistance information, and/or measurement to aid AI/ML operation, where such information acts as a side information and affects the AI/ML model output. For example, such information may be directly fed as an auxiliary input to the AI/ML model, or such information may affect which one among parallel AI/ML models is selected.

2) Joint ML operation between UEs and gNBs: Output from the AI/ML inference from one node is conveyed to the other node such that it may be used as an input to the other node's AI/ML.

For training, though "training categories 0, 1, and 2" could be defined in a similar manner, we are not keen on doing it. Rather, it may be sufficient to say that the following collaboration frameworks may be applicable:

- **Providing training data for online training**
- **Model upload and download**

31 – InterDigital

Distinction between 0b and 1 should be clarified. In addition, there may be sub categories within 1 and 2 considering different collaboration options other than model exchange e.g. model configuration etc., We think that details on categorization can be discussed during the SI.

2.3 General: Evaluation methodology

Based on the discussions in [2], the project on AI/ML for Air Interface should use the following evaluation methodology assumptions:

- Base the evaluation methodology in AI/ML for Air-Interface study on existing 3GPP framework for evaluations, i.e., statistical channel models (from TR 38.901 and TR 38.857 [positioning]), link and system level simulations, etc.
- Additionally, field data to further assess the performance and robustness in real-world environments should be used. How the field data set is obtained (per company vs. common data set) needs to be further discussed.
- Evaluated AI/ML schemes should be based on offline training for evaluation.

Company comments on nomenclature are welcome in the following review form:

Feedback Form 3: Evaluation Methodology

1 – vivo Mobile Communication Co.

We are fine to use 38.901 as the baseline.

For the field data to assess the performance, we prefer to clarify: How the field data set is obtained (per company vs. common data set), **how to evaluate the quality of the data, and how the data is used in combination with the baseline dataset to assess performance** needs to be further discussed.

2 – NTT DOCOMO INC.

Performing the evaluation with field data on top of data generated from channel models would be pragmatic. However, we are concerned about the validity of field data. Accordingly, we prefer discussing how to obtain the field data, before making decision to use field data for the evaluation.

3 – LG Electronics Inc.

We wonder the third bullet would mean that we assume well-trained AI only for evaluation purpose. During SI phase, the impacts of AI training should also be analyzed considering different convergence level when the wireless channel environment changes, e.g. when UE enters a new cell/TRP, blockage scenario, etc.

4 – CAICT

We support statistical channel models, i.e. 38.901 and 38.857, is used as baseline for evaluation methodology. Offline training AI/ML models is used for evaluation.

The usage of field test data should be discussed further. If we preclude that filed test data is included during the evaluation phase, field test dataset is required for the evaluation phase. It is more likely that the way of field test dataset obtaining should be part of the study. Furthermore, how to evaluate the performance of AI/ML schemes with field test data should also be studied.

5 – ZTE Corporation

Generally agree with these principles.

- Utilization of existing 3GPP framework is mandatory for evaluation. Utilization of data obtained from field should be optionally to companies.
- The used AI models should be reported by companies. One reference model for each use case can be aligned among companies at least for calibration.

6 – Guangdong OPPO Mobile Telecom.

We agree to use statistical channel models (from TR 38.901 and TR 38.857 [positioning]) as the baseline channel to evaluate the AI/ML approaches. We have concern on how to evaluate the quality of field data in this study and whether we have enough time budget to do so in Rel-18.

For the offline training, for a given use case the data set used for evaluation (for training, validation and testing) need to be aligned among different proposed AI/ML algorithms

7 – SHARP Corporation

We are fine with moderator's proposal.

8 – FGI

We are generally fine with the moderator's proposal. Regarding the field data, in addition to how the filed data set is obtained, how to evaluate the quality of field data is suggested to consider.

9 – CATT

We support to use statistical channel models defined by 3GPP as baseline for evaluation.

Field data set is essential to access the performance of AI/ML model for practical deployment. The methodology of constructing field data set can be part of the study, including how to evaluate the quality of field data set.

10 – Sony Group Corporation

The comments we make are on the evaluation methodology. We assume the text in 2.3 on "company comments on *nomenclature*…" is a typo.

We generally agree with the assumptions.

We see both advantages and difficulties in using field data, but these issues can be discussed. A common set of field data is preferable in order to allow comparison of evaluations between companies.

Some use cases may benefit from online training, including federated learning. The SI shouldn't restrict itself to offline training. The phrase "based on offline training for evaluation" needs clarification. Does this phrase mean that an implementation could use either online training or offline training but that the evaluation in the study item would only consider offline training. If that is the case, how would the study draw conclusions on the merits of / problems with online training?

11 – Rakuten Mobile

For second bullet about field data, it is essentially needed to be discussed for the method of the field data collection, and the validation of the data should be studied.

12 – HUAWEI TECHNOLOGIES Co. Ltd.

We are generally fine with the proposal. However, one question for clarification is on the definition of offline training here, e.g. it is not clear either input generation or model training is offline, or both are offline. It seems different companies may have different interpretation/understanding, therefore better to discuss this aspect more to align the understanding. In our understanding, offline training here means both input generation and model training are offline.

13 – DENSO CORPORATION

Agree with moderator's proposal.

14 – Samsung Electronics Co.

These evaluation methodology assumptions are a good starting point.

In addition, we may want to study potential updates to the existing 3GPP framework in order to support AI/ML-based evaluations. For example, if a single trained AI/ML model is integrated into a simulation framework, can we guarantee that the simulation results will be repeatable? What if multiple trained AI/ML models (e.g., one per UE) are integrated into that framework?

We may also want to consider the utility of online training for certain use cases (e.g., beam management using reinforcement learning, etc.).

At least for the purpose of feasibility study in the 3GPP SI, including the performance evaluation of AI-powered scheme, and comparison between different typical AI neural network structures, we see the benefits from a common dataset or dataset individually derived from a commonly used methodology (i.e., statistical channel model, link/system-level simulation).

We agree that having field data can be helpful. However, a more detailed level of discussion will be needed. For instance, what kind of field data, e.g., channel measurement, or KPI-related data, would be needed for this RAN1 AI/ML SI? If a data set is obtained and maintained by each company, how to address unfairness between companies, i.e., some companies may have no access to field data. If the common data set is maintained, does 3GPP intend to host a repository? But of course, the filed data provided by companies shall be welcomed, which is potentially useful to be used to verify AI algorithm generalization performance.

15 – BT plc

Field data-based assessments would be desirable from our perspective. We support earlier comments here on evaluation of the type and quality of such data.

16 – MediaTek Inc.

The use of field data should be clarified, our understanding that it'd be used to verify the evaluation based on SLS data as a sanity check. However, this may not available or applicable to all use cases, therefore, we'd suggest to make it optional at least.

17 – Lenovo (Beijing) Ltd

For the evaluation methodology, there are three aspects to be considered: data set construction as indicated by Moderator, AI/ML model selection and AI/ML model generation.

We have the following views on these three issues:

- Data set construction
  For the use cases related with the propagation channels, the *channel models defined in 3GPP can be used* to generate the data set, and we don't need to spend much time during the SI on how the real data sets can be obtained, because it is not realistic to assume the wide availability of real-world data sets for all companies in a timely manner.
- AI/ML model selection
  The basic architecture of an AI/ML model is more meaningful than the optimization on the model in this study to determine the evaluation methodology.
- AI/ML model generation
  The model can be generated locally or downloaded from outside, and possibly tuned and selected via signaling, which is hard to be evaluated. Thus, depending on which of the above categories in Section 2.2 are applied, the evaluation methodology might be needed for the generation methods, such as the overhead and latency in case of model training and transfer in Cat 3.

18 – NEC Corporation

We agree with the proposal in principle.

For the field data, we have concerns for the availability, validity and reliability. It may cause endless discussion on whether the model can be used or not.

To obtain traditional performance metrics like throughput and BLER through SLS/LLS is very time-consuming and can be impacted by many other factors. To save time and to have a reasonable simulation size, we suggest to consider intermediate results as the performance metrics case by case, for example, similarity/MSE of inferenced CSI and perfect CSI for CSI feedback use case.

19 – Fujitsu Limited

We are fine to use datasets generated from statistical channel models referring to TR 38.901 and TR 38.857 for AL/ML training and inference. The evaluation methodology in details should be studied. Regarding to field data, we think it is necessary to study the feasibility and methodology on how to use field data to evaluate AI/ML first. For example, how many scenarios/configurations/parameters from real world is enough to train and/or test AL/ML and confirm its robustness?

20 – Keysight Technologies UK Ltd

We are Ok with 38.901 and 38.857 as a baseline.

We believe that the field dataset to assess the performance should be common for all companies. The SI methodology assumptions should not leave this aspect open for further discussion.

We would like to suggest that one of the study item's goal is to agree on field dataset assumption: company specific or common. We should not start evaluating against field dataset without having answered this question first.

We will need to study synthetic and field datasets for:

1. Training – learning process & fit model models
2. Validation – tune AI/ML models
3. Testing – assess performance of AI/ML models

Another aspect of the evaluation methodology that needs to be considered, especially for field datasets, is what and how the privacy sensitive information can be removed from the datasets gathered from the field either through anonymization, synthesis, or other methods.

21 – Nokia Corporation

We agree that the existing 3GPP frame work for evaluations should be utilized. However, we see that extensions to the current 3GPP NR simulation methodology is needed to enable benchmarking of AI/ML based RAN enhancements, including assumptions for training/exploration and introduction of potential AI/ML-specific KPIs that makes it possible to monitor convergence, stability and accuracy of such schemes. We also see that it is necessary to study and identify applicable AI/ML framework, training requirements, overhead of training including comparison of offline vs online training, complexity of training and deployment. We also see it important to study how to obtain comparability of results including stability and reliability and simulator calibration/repeatability.

22 – ROBERT BOSCH GmbH

In general, we support using a statistical channel models, i.e. 38.901 and 38.857, as baseline. Field data in general for further evaluation on top would be desirable. But we prefer to strive for a validated dataset framework for all use cases which is agreed and supported by all companies.

23 – VODAFONE Group Plc

We don't understand the comments on "validity" of field data (perhaps the people are meaning whether the data is 'representative'?), but, whatever, the important thing is to have some way to check the validity of the statistical models!

24 – Xiaomi Communications

Generally, we are fine with the proposal. As for the field data, we share similar view with vivo/oppo/CAICT that careful study is needed.

25 – China Mobile Com. Corporation

Although field data set is helpful to access the performance of AI/ML model, but the feasibility of using field data is questionable. We think simulation data based on 3GPP channel model can be used, and a common data set can be provided.

Regarding the evaluation methodology of AI/ML for Air-Interface study, we think we need to touch somehow the AI model and algorithm. To make sure the performance comparison is fair, a typical model for calibration is necessary. For performance comparison of different AL models, the used AI/ML models should be reported by companies for crossing check.

26 – Deutsche Telekom AG

The listed items with respect to evaluation methodologies are generally fine with us. Following comments:

- Link/system level simulations based on statistical channel models are well known and allow calibration before starting dedicated evaluations.
- Using in addition field data is certainly a useful further step to analyze AI/ML model behavior in a real world environment, but to allow comparable results the provisioning of common data sets for evaluations purposes covering different scenarios is a basic requirement.
- The requirement on offline trained models for use in Model Inference function is based on a requirement set in RAN3's SI. But it doesn't only require training, but also validation and testing. This should be clarified in the third bullet point.

27 – Fraunhofer IIS

For the evaluation, statistical channel models (38.901) can be used as a baseline in a first step. However, these models are partly over-simplified and they do not always accurately reflect real-world channels. So, it should be avoided to design AI/ML networks that work nicely with the 3GPP channel models, but fail when used with real-world channels. So, the usage of field data seems really necessary for the evaluation phase of this study and should be discussed further.

28 – Intel Technology India Pvt Ltd

We think its beneficial to define common training and testing data sets for comparison of the solutions developed as part of the SI, but it also ok for us to have such discussions in RAN1

29 – Apple Hungary Kft.

Based on statistical channel model, link/system simulation has no issue. Acquiring well validated real world environments data has always been challenging in the world of AI for communication networks.

30 – Ericsson LM

We support basing the evaluation methodology applied during standardization for each use case on an appropriate existing 3GPP framework (e.g., statistical channel models, link level simulations, system level simulations).

We support companies sharing results and conclusions drawn from field data (e.g., to verify ML model behavior, robustness, and performance).

We do not support sharing field data (the need to share field data has not been justified from our perspective).

31 – Futurewei

• We are ok to start with using existing 3GPP framework for evaluation.

• Using field data is reasonable when it's available and carefully selected. Depending on the use case, it may not be feasible for some use cases. Thus, we suggest making it clear that this should be use case-dependent and using field data only when it's feasible and clearly specified.

• For the 3rd bullet, even though evaluating AI/ML methods based on offline training is less complicated, we suggest not adding such restriction at pre-SI stage / before use cases being agreed.

32 – Qualcomm Incorporated

We support the proposal.

33 – InterDigital

We support using statistical channel models (from TR 38.901 and TR 38.857 [positioning]) for evaluation. As pointed out by other companies, we have concerns regarding additional efforts to define, collect and validate representative field dataset.

2.4 Use Cases for AI/ML for Air Interface

Use cases for the candidate project on AI/ML for Air Interface were heavily debated in the previous email discussion [2]. A number of companies mentioned the possibility of discussing/deciding the use cases for this study during the first phase of the project, while other companies were willing to agree on a limited set of promising use cases before the start of the project.

It is the Moderator's understanding that the use cases for the study should be representative of potential areas where the application of AI/ML based techniques are expected to render benefits over traditional techniques. The selection of a number of representative use cases should not be seen as exclusive but rather as a way to identify a meaningful 3GPP framework for AI/ML for Air Interface which could be extended to other use cases in the future, e.g., in relation to subsequent potential WIs.

Based on the popularity of use cases discussed in [2], the Moderator's proposal is to focus the 3GPP framework for AI/ML for Air Interface on:

- CSI feedback compression (lower overhead)
- Beam management (beam selection, beam recovery, …)
- Positioning

Company comments on Moderator's proposal on use cases. If you disagree with the proposal, please explain why you think the above three use cases would not sufficiently exercise the 3GPP framework for AI/ML for Air Interface.

Feedback Form 4: Use cases

1 – vivo Mobile Communication Co.

We support all the three use cases.

In addition, we also believe RS overhead reduction, channel estimation performance enhancement or mobility performance enhancement could benefit from application of AI/ML. Including these use cases is beneficial at least for the following aspects:

1. This would provide a whole picture for what would be possible within 5G framework and provide some initial input for future evolution in the next 2 or 3 releases.

2. Some of the low hanging specification enhancement are expected for above mentioned use cases.

3. For selection of use cases, another aspect that need to consider is the efforts for evaluation including construction of data set for each use cases. The above listed use case (e.g., RS overhead reduction) could easily provide large amount of data samples for training and verification compared to the listed use cases from the moderator.

Even with limited set of use cases, we would like to point out that a pre-study phase is still needed due to the following reasons:

1. for each of the use cases, the intended target categories from different companies and the related evaluation scenarios for different categories would be drastically different. A pre-study phase would help the group to have clear understanding before directly starting the evaluations.

2. 3GPP is not familiar with AI/ML related background information, including typical models used for different cases, detailed training and inference procedures etc. The group needs a pre-study phase to collect more information for a well-prepared start.

2 – NTT DOCOMO INC.

We agree with deciding representative use cases to reduce workload for study item. As an evaluation methodology depends on AI/ML use case, it might be necessary to set up as many evaluation methodologies as AI/ML use cases to study. We believe having moderate number of representative use cases leads to success in AI/ML for air interface in Rel-18.

In the lists of use cases, we prefer renaming "CSI feedback compression" to just "low overhead CSI feedback", because we are not sure if companies have the same understanding about the CSI feedback compression definition. Our preference is not to apply AI/ML compression scheme of CSI feedback. Instead, AI/ML is deployed only on NW to reconstruct CSI with sparse CSI feedback as an input.

We are fine with other two use cases to be studied in Rel-18.

3 – LG Electronics Inc.

Despite moderator's explanation, we wonder whether this proposal precludes including study results of other use cases than the listed three. We think that the first phase of SI addresses general issues for AI for air-interface, e.g. evaluation assumption, categorization, AI model, etc. During this phase, it may be fine to focus on the listed three use cases. In the second phase, we can proceed studying use cases and potential

benefits of each use case. For the second phase, we think that it is too early to limit the SI scope only for the listed three use cases. Especially, RRM/mobility enhancement would have different aspects and can be discussed in RAN2 in the second phase while other use cases can be handled by RAN1.

4 – TELECOM ITALIA S.p.A.

The scope of this study seems very large. Suggest to focus on one or max two use cases, with priority on beam management

5 – CAICT

We support study the listed 3 use cases. Besides, RS overhead reduction, mobility enhancements should also be included. A use case study phase is preferred to have a clear understanding of each use cases and potential use case selection for further evaluation in R18.

6 – Spreadtrum Communications

We prefer to keep the five use cases listed by moderator in RAN#93e during the final round discussion on NWM, which seems to have achieved the majority support. We should set up one common basis for all use cases as possibly in Rel-18, e.g., data set construction, AI model, evaluation methodology and so on. We can further down-select use cases during SI stage.

7 – ZTE Corporation

We are supportive to have these three use cases for Rel-18 study. We think it is a good balance considering time limit and companies' interest. The following includes our further suggestions on specifying each use case in the SID.

- CSI feedback compression: Study channel matrix or precoding matrix compression feedback in at least spatial domain and frequency domain
  - Use Rel-16 eType II codebook as the baseline scheme
- Beam management: Study beam prediction in time or spatial domain to improve UE experience in FR2 with reduced the beam measurement RS overhead and UE measurement/reporting
  - For beam prediction in time domain, focus on high mobility scenario (e.g., high-speed train and high-way)
  - Study aspects include evaluate the feasibility and potential system level gain and specification enhancements such as beam measurement, beam report, beam indication and fallback approach
- Study positioning accuracy enhancement in scenarios with high NLOS probability, e.g., IIOT DH

8 – AT&T

We agree with NTT DOCOMO that "CSI feedback compression" could be renamed as "low overhead CSI feedback"

In addition, we support adding the use case of "Link Adaptation and HARQ feedback." The configuration of parameters related to link adaptation have to balance multiple objectives including coverage targets, spectral efficiency, QoS/reliability requirements for some traffic types (e.g. URLLC, V2X, XR, etc.) and also involve balancing complex interactions between gNB and UE capabilities and varying implementations. ML-based techniques can be applied to both assist in the adaptation of configurations to optimize

certain air interface KPIs but may also be used to develop more efficient feedback and resource allocation signaling approaches as well. CSI feedback is certainly related, but is just a specific subset of a larger link adaptation framework which can benefit from tightly integrated ML algorithms in the RAN.

9 – Guangdong OPPO Mobile Telecom.

We are fine in general to confine to 3 use cases. But we prefer RS reduction to beam management. In limited simulation on the table it seems the study on beam management usually are based on ideal assumption e.g. low speed UE hence the study on beam management may not help to improve the performance in system level but some limited scenarios.

10 – SHARP Corporation

We are fine with the listed three use cases. However, we are not sure it is reasonable to preclude the other two use cases (RS overhead reduction and mobility). Further discussions are needed to reach a certain level of consensus.

We share the same view from Docomo that "low overhead CSI feed- back" could be more suitable, which covers both CSI feedback compression (reference performed at both UE and gNB) and sparse CSI feedback (reference performed only at gNB).

11 – CATT

We prefer to limit the number of use cases in the detailed study, i.e., we should only conduct performance evaluation for limited number (1-2) of use cases. If there is a pre-study phase, more use cases can be studied in the pre-study phase. As an output of the pre-study phase, limited number of use cases for further study can be selected.

We're ok the use case of CSI feedback compression and positioning. Beam information can be considered as a kind of CSI. Beam management use case is not very essential if there is already a use case for CSI feedback.

12 – Sony Group Corporation

We think use cases shouldn't be down-selected at this stage as other companies also mentioned in last email discussion. We should have an initial phase of the study to identify use cases, encouraging new use cases beside the afore-mentioned three use cases.

An issue with considering a single use case is that the use case may not illustrate the full range of issues with supporting AI/ML in the air interface. We need to consider several use cases to work out what are the potential impacts on the air interface. For example, if the study just consider the CSI operation with sparse CSI feedback, then the study could conclude that AI/ML could be implemented solely in the gNB without additional interaction from the UE. Such a conclusion would not be generally applicable, for example autoencoders may require significant signalling between the UE and gNB, even to the extent of model transfer.

In summary, the AI/ML study needs to consider a sufficient number of use cases to make some general conclusions on how AI/ML could impact the air interface.

We don't support the phrasing of the question "*explain why you think the above three use cases would not sufficiently exercise the 3GPP framework for AI/ML for Air Interface*". This seems to pre-suppose that the purpose of the study is to look at the *framework* for AI/ML. We take a broader view of the study where studying the framework is only one aspect of the study. The study additionally includes use cases, evaluations, training data, specification impacts etc.

13 – Rakuten Mobile

We share the opinion with LG's proposal that it is too early to preclude other use cases, and thus it requires more discussions for use cases at this time.

14 – HUAWEI TECHNOLOGIES Co. Ltd.

"CSI feedback compression (lower overhead)" should be revised to "CSI feedback enhancement". It is well acknowledged that one of the sustained issue for CSI feedback mechanism in practical networks is the accuracy of CSI for the real-time channel condition. Overhead reduction for CSI feedback can be one of the aspects to be studied also.

In addition, we also think RS enhancement is a good use case for AI/ML for air-interface. RS can be enhanced with AI/ML by either reducing the overhead or increasing the number of antenna ports, which is important for overall system performance.

15 – DENSO CORPORATION

We also agree that the proposed three use cases are viable. In addition, mobility is also a potential area to study if the performance can be enhanced by AI/ML.

16 – Samsung Electronics Co.

Based on the moderator's statement, "The selection of a number of representative use cases should not be seen as exclusive but rather as a way to identify a meaningful 3GPP framework for AI/ML for Air Interface which could be extended to other use cases in the future, e.g., in relation to subsequent potential WIs.", these three use cases are generally fine. That is, potential discussions on other use cases should not be precluded. On the other hand, 'CSI feedback compression' can be generalized to 'CSI feedback' to include other relevant topics such as future CSI prediction. RS overhead reduction is also worth to be considered.

Considering it is the first SI on AI/ML air interface, at least two different usages of AI/ML need to be considered: online trained or offline trained AI./ML. It is suggested to at least have one use case for both online and offline trained AI/ML.

For the topic of Positioning, the current description is too vague, and it is not clarified that what particular aspects are intended to be improved by AI/ML.

17 – BT plc

The themes of the use cases look reasonable, but would like the use cases to be closer linked to service perfromance.

For example, while lowering CSI overhead is generally desirable, it in itself is use-case dependent, i.e is not an end goal. A capacity/coverage/reliability-related enhancement benefitting from AI/ML-driven lower CSI overhead should to be part of the use case.

So we prefer to change "CSI feedback compression (lower overhead)" to "low overhead CSI feedback for capacity and coverage enhancement", and if possible have a similar linkage to some service objective for any other identified use cases.

18 – Lenovo (Beijing) Ltd

We prefer to have the two use-cases of CSI feedback compression and beam-management, because the more use-cases mean more efforts needed to do evaluation.

Furthermore, we also think that for each of the use cases, different categories need to be evaluated as listed in Section 2.2, and based on the evaluation, it can be determined that which category would suffice for which kind of AI/ML-based approaches.

19 – MediaTek Inc.

First, the scope of Positioning should be clarified, for example, use AI/ML to improve positioning accuracy, or for some other purpose.

Second, we understand that the goal of such R-18 study is an initiative to evaluate the benefit of AI/ML integration with example air interface functions in each WG. However, from the listed use cases, no R2 function is listed, which will result in no study in R2.

20 – NEC Corporation

We agree with the 3 cases. In addition, we also think RS overhead reduction and mobility should be included. These two cases were supported by many companies in the September email discussion. So it will be useful to study them in the SI phase. For the "pre-study" phase for the use cases, do we need to list it as a discussion topic and collect companies view on it?

21 – Fujitsu Limited

It is fine to have several representative use cases as the start point of the entire SI stage. But it might be too early close the door to other than the listed 3 use cases. We can use these 3 use cases to clarify the common understanding among companies on AI/ML framework, evaluation methodology etc. The decision on Rel-18 use cases finally can be done later.

22 – FGI

We are fine with the three representative use cases to be a start point and serve as examples to help us study AI/ML for air interface. However, we cannot preclude other use cases since they may also benefit from the study outcome of AI/ML for air interface.

23 – Keysight Technologies UK Ltd

We agree with the moderator's proposal for use cases.

Although there are other interesting use cases like RS overhead reduction. We need to be careful when scoping the SI. As there are many other questions that need to be answered (e.g., KPIs, evaluation methodology, etc.), we believe we should not try to have a too broad set of use cases.

24 – Nokia Corporation

In our view use cases should be started as part of the study. So far no technical analyses have been presented why these proposed use cases would be presentative for the AI/ML studies. The proposed use cases could be considered as starting point for the studies but further technical studies are needed before use cases can be decided. Furthermore, at least point we should not try to list sub-options for the initial list of use cases since no technical studies have been conducted for the proposed use cases and related details. Even the

main use case bullets without further studies and descriptions are not clear and can be understood rather differently by different companies.

25 – ROBERT BOSCH GmbH

We support all the use cases mentioned by the moderator but propose to consider mobility performance enhancement in addition.

26 – VODAFONE Group Plc

We see CSI enhancement (e.g. CSI estimation and prediction) as the priority use case, and, Beam management (beam selection, beam recovery, …) as another useful use case. Other use cases may also be good to study, e.g. mobility.

27 – Xiaomi Communications

We could understand it is benefical to limit the use cases considering limited time. But on the other hand, which use cases should be selected should be based on some technique discussion . In our understanding, there is no comprehensive technique discussion and analysis for each use case. Thus, we think it is better not to preclude RS overhead reduction and mobility at this stage. A pre-study phase can to set to identify the use case and then companies could have better understanding in all the five use cases.

28 – China Mobile Com. Corporation

In our views, one or two of the listed use cases can be selected to conduct performance evaluation. In Rel-18, we should focus on the evaluation methodology. More use cases can be considered in future releases based on the identified evaluation methodology.

29 – Deutsche Telekom AG

We are fine with the 3 proposed use cases, but as some other companies we wouldn't exclude already now the other 2 use cases listed in [2]. With 5 use cases under consideration a broader view on AI/ML principles and methods would be feasible.

30 – Fraunhofer IIS

In our view, the three use cases are reasonable for the R18 AI/ML study. The scope of each use case needs to be clarified (e.g., what is the objective of the positioning use case? Is it to use AI/ML to improve accuracy, or something else?).

31 – Intel Technology India Pvt Ltd

Support. We believe down-selection is beneficial to focus RAN1 scope, a use-case study in RAN1 is not an efficient usage of RAN1 TU. Propose more details as below:

- CSI: Precoder determination in spatial, frequency and time-domain with reduced overhead
- Beam-management: Beam determination with reduced latency and possible measurement/reporting overhead reduction
- Positioning: Positioning performance enhancement in NLOS conditions

32 – Apple Hungary Kft. We agree in general to limit the scope of the SI. Changing “CSI feedback compression” to “ CSI feedback(low feedback overhead, high CSI feedback fidelity, etc)” for the tradeoff between overhead & fidelity.
33 – Ericsson LM We support CSI, beam management, and positioning as representative use cases. For CSI, we think the scope should expand to cover both lower overhead (or increased spectral efficiency with existing CSI overheads) and resilience against UE mobility. We also agree: The selection of certain representative “pilot” use cases is not exclusive. We see this SI as a means to identify a meaningful 3GPP framework for AI/ML for Air Interface that could be extended to other use cases in the future (e.g., in relation to subsequent potential WIs). We support limiting the number of use cases to at most three.
34 – Futurewei • We support the proposed 3 use cases. However, companies seem to have different views on the ”CSI feedback compression (lower overhead)” use case, thus, we suggest the moderator providing some clarification after considering views across companies. • During last email discussion, we observed “RS overhead reduction” or ”RS enhancement” use case also gained support from many companies, thus, we suggest including it in the use case list.
35 – Qualcomm Incorporated **We support the Moderator’s proposal.**
36 – InterDigital For a first potential study item on AI/ML for Air interface it is important to limit the number of use cases. We agree that CSI feedback compression use case can be generalized to CSI feedback enhancement.

2.5 KPIs

Previous discussions as captured in [2] led to the following:

- KPIs are broadly understood to be use case specific but a number of common metrics prevail, e.g., complexity and performance vs. proper (non-AI/ML) state-of-the-art baseline. Similarly, overhead associated with enabling respective AI/ML scheme should be well documented and accounted for.

Company comments on KPIs are welcome in the following review form:

Feedback Form 5: KPIs

1 – vivo Mobile Communication Co. We share similar understanding as the moderator.
2 – NTT DOCOMO INC. We agree that KPIs are broadly use case specific and some KPIs might be common across different use cases.
3 – LG Electronics Inc. We agree that KPIs are use case specific and overhead needs to be well documented and accounted for.
4 – CAICT Support moderator's proposal.
5 – Spreadtrum Communications Support Moderator's proposal
6 – ZTE Corporation General agree to use typical KPIs in 3GPP for LLS and SLS, e.g., throughput, BLER, MSE, which should be use case specific.
7 – AT&T Support the moderator's proposal.
8 – Guangdong OPPO Mobile Telecom. We agree in general KIP is use case specific. Then we are wondering what the overhead is in this context?
9 – SHARP Corporation We agree with moderator's proposal.
10 – FGI We support the moderator's proposal.
11 – CATT We agree that overhead associated with enabling the AI/ML algorithm shall be considered.
12 – Sony Group Corporation As moderator mentioned, this is related to the use cases, so we should discuss the details of KPIs and evaluation methodology for each use case in a new study item of AI/ML for air interface. The KPIs should include the associated signalling and procedures for supporting the AI/ML algorithms. The KPIs can also consider the reduction in signalling that is possible when an AI/ML approach is taken

(e.g. the study can consider whether / how much mobility signalling is reduced for the mobility enhancements use case).

Hence, we think that there will be a mixture of some use-case specific metrics and some generic metrics, as suggested by the moderator.

13 – Rakuten Mobile

Support moderator's proposal

14 – HUAWEI TECHNOLOGIES Co. Ltd.

We agree with the above description on KPIs. For example, complexity is one of the key KPI for model training and model inference. For model training, complexity here can reflect factors like training complexity and amount of data samples. For model inference, complexity here can reflect factors like computational complexity, e.g. the maximum time complexity (FLOPS) and size of AI/ML models. Companies should also report the time complexity and size of AI/ML models used for evaluation.

15 – DENSO CORPORATION

Share the same view as moderator.

16 – Samsung Electronics Co.

No concerns with this summary. Although there could be a few KPIs that can be commonly applicable, it is clear that performance KPIs will be used case-specific.

Different from performance KPI, the complexity KPI for the inference phase can be more complicated in multiple dimensions, including memory occupation, computation complexity, and latency, regression or classification accuracy, etc.

17 – BT plc

We agree with the summary.

18 – Lenovo (Beijing) Ltd

As presented by Moderator, there are two kinds of KPIs in this SI to be identified, one is the traditional KPIs for the selected use cases, and the other is dedicated KPIs for the AI/ML approaches. We have the following comments:

- Traditional KPIs for the selected use cases, for example,
  - For CSI feedback: CSI-RS resource overhead, UCI overhead, system performance loss/improvement (e.g., throughput in LLS, spectrum efficiency in SLS), NMSE of the recovered CSI
  - For beam management (BM): beam acquisition overhead and latency (e.g., number of CSI-RS for BM), received SINR/RSRP with the selected beam pair, system performance loss/improvement (e.g., throughput in LLS, spectrum efficiency in SLS).
- Dedicated KPIs for AI/ML approaches, we think the following aspects should be considered:
  - Measurable: To evaluate model generation, the latency and frequency of required model for Cat.3 should be considered, which are applicable to all use-cases.
  - Quantifiable: The complexity of the inference operations should be quantified, such as the number of OPs and/or latency.

- Comparable: The non-AI/ML approach should be also evaluated with the same metric as the AI/ML-based approach on the complexity of inference.

19 – NEC Corporation

We also agree with the moderator. We want to confirm if the 'complexity' also includes the energy consumption as the training process may cause non-negligible energy consumption.

20 – Fujitsu Limited

we support moderator's proposal.

21 – Keysight Technologies UK Ltd

We agree with the moderator's point of view. Although there are generic AI/ML-specific KPIs that should also be considered as they can be used as inputs for evaluating the models' performance, e.g., training accuracy, validation dataset accuracy, bias, and variance.

22 – Nokia Corporation

In addition to the KPIs listed in moderator's summary we see that the following two metrics should be considered in the studies of different use cases; 1) comparability of simulation results including stability and reliability and simulator calibration/repeatability and how to 2) possibility to define UE requirements and verify UE requirements with test cases. It would also be important that companies have common understanding on what parameters, set of parameters or assumptions to use for training data generations. It is important to show system performance gains for AI/ML, not only e.g. UE power saving gains or UE implementation complexity reductions. Furthermore, it is also important that these performance gains are available in the UE requirements and can be verified with test cases.

23 – ROBERT BOSCH GmbH

Agreed.

24 – VODAFONE Group Plc

OK, but probably need to separate the training and inference KPIs/measurements

25 – Xiaomi Communications

We agree with proposals. Besides, the power consumption should be set as one KPI and a power consumption model can be studied to evaluate the power consumption performance during training/ inference.

26 – China Mobile Com. Corporation

The KPIs should include two kinds of metrics. One kind is traditional metrics for LLS and SLS, e.g., throughput, BLER. The other one is metrics for AI model, including complexity, reliability/generalization.

27 – Deutsche Telekom AG

We generally agree with moderator's view.

Besides system performance related KPIs it is very important to evaluate the additional processing effort and data overhead an AI/ML scheme consumes against the baseline w/o AI/ML usage. It is further required

to differentiate between initial offline training efforts to create/parametrize the model and the efforts during run time in Model Inference function (eventually together with online training).

28 – Intel Technology India Pvt Ltd

The following KPI could be added as part of common metrics: sensitivity of the AI/ML scheme performance to possible model mismatch compared to non-AI/ML state-of-the-art baseline. Example: data-set used for training CSI model is mainly from outdoor environment (due to data-collection restrictions from indoor residential scenario) but same model is also used indoors.

29 – Apple Hungary Kft.

Agree with modulator's understanding

30 – Ericsson LM

We support the proposal.

31 – Futurewei

• In general, we agree with the above identified example evaluation metrics for AI/ML-based approaches, i.e., performance, algorithm complexity, and overhead.

• In the cases that the solution consists of AI/ML and non-AI/ML components, we suggest evaluating AI/ML model performance separately from use case solution performance.

32 – Qualcomm Incorporated

We support the Moderator's proposal.

33 – InterDigital

We agree with broad classification of use case specific KPIs and common KPIs. Common KPIs can include model complexity, model size, inference latency, power consumption etc..

2.6 3GPP framework for AI/ML for Air Interface

One of the allegedly main objectives for this SI would be to define the 3GPP framework for AI/ML for Air Interface. Would you agree? If so, how would you think the corresponding objective could be stated in the candidate SI.

Company comments on 3GPP framework for AI/ML for Air Interface are welcome in the following review form:

Feedback Form 6: 3GPP framework for AI/ML for Air Interface

1 – vivo Mobile Communication Co.

We have the following objectives in mind in defining the 3GPP framework for AI/ML for Air interface:

for each of the categories identified in pre-Study phase, study the framework and procedures of applying AI/ML for the target use cases, including potential specification impact and related specification effort in

other TSG; [RAN1/RAN2/RAN4]

2 – NTT DOCOMO INC.

We agree.

It could be stated as "Study the 3GPP framework for AI/ML for air interface corresponding to each target use case from the aspects of performance, complexity and potential specification impacts."

3 – LG Electronics Inc.

We agree that the main target of SI would be to define the 3GPP framework for AI/ML for air interface. SI objectives can include AI/ML categorization, evaluation assumptions/KPIs, and analysis/evaluation results showing potential performance of AI/ML based approaches. The study results can be used for scoping subsequent WI scope, possibly in Rel-19. As suggested in the answer for #4, we prefer to split SI into two phases. In the first phase, non-use case specific issues are handled such as categorization and evaluation assumption. In the second phase, use-case specific issues are handled, which targets performance analysis of AI based approach for each use case. In the second phase, RAN1 and RAN2 may proceed in parallel with separated set of use cases, e.g. BM/CSI in RAN1 and RRM/mobility in RAN2.

4 – CAICT

We agree that this SI is to define the frame work for AI/ML for Air Interface. Generally, we observe that the enhancements by AI/ML for air interface is use case specific and the collaboration framework provides a guidance for the potential specification works. It could be considered that collaboration framework is included as part of the study.

Some other common understandings, like detailed AI/ML model should be left for implementation and will not be specified, could be captured in SI as note(s).

5 – ZTE Corporation

Agree to study 3GPP framework for AI/ML for Air Interface. We can generally say

- Study potential specification enhancements in terms of Cat 0b), Cat 1) and/or Cat 2) for the identified use cases [RAN1, RAN2]

For each of the three use cases, suggest to make the description clear and specific as we comment in 2.4. This is to make sure we have a clear target for Rel-18 study. The detailed description of Cat 0b), Cat 1) and Cat 2) subjects to the outcome of the discussion in 2.2.

6 – Guangdong OPPO Mobile Telecom.

We think apart from spec impact, we need take both pain and gain into account as a whole. So apart from performance, we think TR should also capture the requirement on storage, computing power and overhead over the air etc. into account, especially from UE point of view

7 – SHARP Corporation

We agree that a main target of this SI is to define the framework for AI/ML for Air Interface. The framework should be supported by the 3 5 use cases that will be studied case by case in details. At the same time, the generality of framework should be maintained such that the framework could further provide guidance/assistance to other use cases.

8 – CATT

We agree that the study is to setup 3GPP framework for AI/ML for air interface. That includes evaluation methodology, AL/ML categorization, and potential signaling/procedures to enable AI/ML.

9 – Sony Group Corporation

We think the objective can be rather open-ended, given that this is a study item. We would be OK with: "study the 3GPP framework for AI / ML for the air interface". The study can then determine what constitutes a "framework". Our initial view is that the framework could include the signalling needed to swap capability information, provide assistance information, aid in inference, aid in training etc.

10 – Rakuten Mobile

We agree to define the 3GPP framework for AI/ML for Air Interface. The implementation method discussion can be postponed till the use cases are defined, and it should be specified based on the use cases.

11 – HUAWEI TECHNOLOGIES Co. Ltd.

In our understanding, the current framework should be taken as baseline for us to account for AI/ML for physical layer. For example, the principles identified during the SI of FS*NR*ENDC*data*collect applies. It should be sufficient to note in the SID as below, no need to add any additional objective for it.

- Detailed AI/ML algorithms and models are left for implementation;
- User data privacy needs to be preserved;
- Support of AL/ML shall reuse the existing RAN architecture and new interfaces shall not be introduced.

In addition, the need to add an objective for framework is not clear to us yet. We are doing study item in Rel-18 for AI/ML, the key goal of the study should mainly focus on evaluating/identifying potential beneficial use case for AI/ML. As to potential signaling/procedure to enable AI/ML for the identified use case, it might belong to WI phase.

12 – Samsung Electronics Co.

Need more clarification on the 3GPP framework for AI/ML for the air interface. Does "3GPP framework for AI/ML for Air Interface" mean "what needs to be done to support AI/ML for Air Interface in a 3GPP-compliant system"? If so, then that sounds like a reasonable objective for this SI. Furthermore, the study on the framework should also include: how to categorize different AI-neural networks to some typical categories, how the generalization performance is to be evaluated, etc.

13 – DENSO CORPORATION

Yes, agree. It is also related to the evaluation and selection of the collaboration framework for the subsequent work in Rel-19. How to state it in the SID could be something like already suggested by vivo and docomo.

14 – Lenovo (Beijing) Ltd

Agree. It should determine the objectives for this SI, and we have the following views:

- Study and identify evaluation methodology on the AI/ML-based approaches [RAN1, RAN4].
    - Determine the use-case specific KPIs and benchmarks of the selected use-cases

- Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels.
- Determine reasonable dataset construction for training, validation and test for the selected use cases.

- Study a general framework to involve the AI/ML capability [RAN1, RAN2]
  - Investigate the potential impacts on the physical/MAC layer procedures, signaling and UE capability, for potential normative work in Rel-19.

15 – MediaTek Inc.

We agree the goal of study is to establish 3GPP framework for AI/ML integrated Air Interface.

16 – NEC Corporation

We agree. We suggest to discuss this topic when other parts of this document have been settled. This could help to decide the scope and wording for the objective discussion.

17 – Fujitsu Limited

We agree that the main objectives of this SI is to define the 3GPP framework for AI/ML for Air Interface. Also, the gain that AI/ML would bring to air interface should be evaluated. At the same time, the feasibility of using AI/ML to air interface should be better concluded as well, such as potential complexity/computing/latency increase. Also, AI/ML's generalization capability and robustness in real-world application need to be clarified through the SI.

18 – FGI

Yes, we agree the goal would be to define the 3GPP framework for AI/ML Air Interface, which may include the feasibility study and performance (e.g., complexity, signaling overhead). In addition to the representative use cases, the study outcome would be the foundation for other use cases.

19 – Keysight Technologies UK Ltd

Although evaluation and discussions would likely be focused on the agreed use cases, in our view, we should strive to obtain a framework from this SI that can incorporate other future use cases. So, in that sense, we should also consider general design aspects so that, when possible, the same questions won't have to be discussed each time a new AI/ML enhanced use case is added in future releases.

If we assume the consensus that models themselves should not be specified and left to implementation. Then, the objectives of the SI should be, for each use case, category, and phase (i.e., training and inference):

- Study specification impact
- Study specification effort
- Study interoperability and testability aspects
- Study privacy aspects of the datasets
- Study a framework to establish and compare models' performance, complexity, and overheads

20 – Nokia Corporation

If we limit the study to few use cases only and without even studying how representative and useful these particular use cases are for studying understanding the implications of AI/ML for air interface, it may be difficult to define the 3GPP framework for AI/ML for air interface based on these studies. Therefore, in our view it would be important to study and identify suitable and representative set of use cases first in order to utilize finding for defining the 3GPP framework for AI/ML for air interface.

21 – ROBERT BOSCH GmbH

1) We agree that the goal is to define a 3GPP framework for AI/ML Air Interface. We think that the objectives should include the evaluation methodology, respective assumptions and the KPIs used for evaluation. Based on a possible categorization of the potential performance of AI/ML approaches should be analyzed and evaluated.

22 – VODAFONE Group Plc

OK. Agree with Sony (#9)

23 – China Mobile Com. Corporation

In our view, the study item should identify the evaluation methodology in AI/ML for Air interface, and make clear that whether AI/ML based method can provide benefit for the target use cases. We have the following views for the SID:

- Study the evaluation methodology of AI based physical layer enhancements
- Study the dataset for each use case, including the source, the size and the structure of the dataset
- Study the evaluation metrics, e.g., throughput, complexity, generalization, etc
- Study the typical AI models for each use case at least for calibration

24 – Deutsche Telekom AG

The question is what is meant by a framework. From our perspective it should cover at least:

- Interrelation (workflow) of functions in a gNB and potentially in a UE to set up AI/ML use case specific enhancements (related also to categorization of AI/ML procedures)
- Use case specific input and output data for Model Inference and Model Training functions (offline/online)
- Description of lifecycle management of AI/ML models (depends on OAM involvement –> SA5)

Note that we don't incorporate here the use cases and the evaluation methodology, which are seen as separate items.

25 – Intel Technology India Pvt Ltd

We think the outcome of SI would define or be used to define such a framework. No need to decide before SI starts

26 – Apple Hungary Kft.

Agree the main objective for this SI would be to define the 3GPP framework for AI/ML for air interface. The framework should include general evaluation methodology, KPI, input/output/location of AI/ML training/referencing, and collaboration framework. Since AI/ML is general use case dependent, the framework should be generic to cover future use cases.

27 – Ericsson LM

The aim for the SI is to develop a standardization framework to introduce AI/ML based features in the physical layer. The SI should also, for some selected initial use cases, provide initial assessment of the benefits relative to current state of the art baseline.

28 – Futurewei

• In general, we agree studying the framework to support "AI/ML for Air Interface" is one of the objectives for this Rel-18 SI.

• As RAN3-led Rel-17 SI "FS*NR*ENDC*data*collect" has defined RAN intelligence framework to support AI/ML-based solutions, we suggest leveraging the framework specified in the Rel-17 SI as a starting point. At the same time, we can identify modifications and/or evolutions that are needed in the framework to support AI/ML use cases for Air Interface.

• FUTUREWEI's view for the objectives of the Rel-18 SI:

a) Framework to support "AI/ML for Air Interface" use cases (using the framework from Rel-17 SI "FS*NR*ENDC*data*collec as a starting point)

b) Evaluation methodologies for identified use cases

c) Standards/specification impact for identified use cases

29 – Qualcomm Incorporated

We agree. R18 should study architecture and procedure enhancements for AI/ML, including

- **Network entities and interfaces for model training/inference, data management, model management (RAN2/RAN3)**
- **Procedures for online training and federated learning**
- **Procedures for inference configuration in UE and network**
- **Procedures for data management and model management**
- **Mobility support for AI/ML based solutions**
- **UE assistance for AI/ML configuration.**

Our understanding is that these will be handled in the RAN2 portion of the AI/ML for Air Interface SI.

30 – InterDigital

The objective of the SI is to study a standardization framework to support AI/ML for Air interface. We think the framework for could also include the means to abstract/characterize UE capabilities w.r.t AI processing.

2.7 Other

Please state any other areas that you think could be relevant for the drafting of a candidate SI for AI/ML for Air Interface:

Feedback Form 7: Other

1 – vivo Mobile Communication Co.

Based on above comments, we would like to provide our thoughts for the objectives in mind:

Phase 1(use case study phase)

- Identify and categorize use cases for AI/ML based air interface enhancements, including at least CSI-feedback, beam management, positioning, reference signal overhead reduction, mobility [RAN1/RAN2]
- Collection of basic information for each use case, including working mechanism, dataset, AI/ML model, UE and network involvement, etc.
- Typical deployment scenario(s) and KPIs for each use case
- Categorization of use cases based on collaboration frameworks between UE and network;
- Down-selection (if needed) of use cases for Phase 2 study

Phase 2(performance evaluation and specification impact study phase)

- Performance evaluation of the use cases concluded in Phase 1 [RAN1/RAN4], including
- Study of evaluation methodology, including dataset construction, model alignment between companies, etc. (Note: conclusion achieved during email discussions phase on simulation methodology could be captured here.)
- Evaluation assumptions and performance comparison for each use case
- Framework and procedures of applying AI/ML for the use cases concluded in Phase 1, including potential specification impact and considering the specification efforts from other TSGs; [RAN1/RAN2/RAN4]

2 – CAICT

We share the same views as vivo listed above.

3 – Samsung Electronics Co.

It may be good to consider lifecycle management of AI/ML models (e.g. data collection, model training, model deployment, model monitoring, model updating). When a UE is involved in this process, they would not be purely upper layer issues.

4 – Keysight Technologies UK Ltd

We share a similar view regarding SI phases as proposed by other companies. The SI should be structured in two defined phases.

- The first phase should be focused on:
  - Completely defining the final agreed uses cases in section 2.4
  - Identifying specific KPIs for the final agreed uses cases in section 2.4
  - Identifying the specific aspects of synthetic dataset generation
  - Evaluating privacy aspects of the datasets
  - Evaluating and deciding whether the field dataset should be common or company-specific for each of the final agreed uses cases in section 2.4

- The second phase should be focused on:
  - Study evaluation methodology and model alignment between companies for each of the final agreed uses cases in section 2.4
  - Performance comparison of each of the final agreed uses cases in section 2.4 using the synthetic and field datasets
  - Study specification impact and effort
  - Study interoperability and testability aspects

Even if, for evaluation purposes in this SI, models are trained offline, we should also consider as part of this SI how the AI/ML model used in a Model Inference function must be initially trained, validated and tested before deployment.

5 – VODAFONE Group Plc

Samsung's comment (#3, above) about UE involvement in 'end of life' of AI model seems important.

6 – Deutsche Telekom AG

The LCM as raised by Samsung is an important topic that has to be covered in the SI (see also our comments to 2.1 and the framework in 2.6).

3 Intermediate Round

The Moderator would like to thank for all the comments received in the Initial Round of discussions.

In the Intermediate Round of discussions, I would like to confirm the candidate project properties. In addition, I would also like to collect feedback on the "Justification" and "Objectives" sections of the candidate project SID.

Those sections have been written as a result of the discussions during the Initial round and with the best Moderator's intentions to progress with the SID definition for this project.

3.1 Candidate Project Properties

Based on the received inputs, the following candidate project properties seem to be agreeable where the note on "no foreseen impact to/from SA" has been removed:

Proposal Int-1: Candidate project characteristics

Table 2: Candidate Project characteristics

Title	SI/WI	Primary WG	Secondary WG	Notes
SI on AI/ML for Air-Interface	SI	RAN1	RAN2, RAN4	SI for entire Rel-18 duration.

Can you agree on Proposal Int-1? Please include your comments on the feedback form 8.

Feedback Form 8:

1 – vivo Mobile Communication Co. Support
2 – NTT DOCOMO INC. We support the proposal.
3 – SHARP Corporation We support proposal Int-1.
4 – VODAFONE Group Plc We support the proposal.
5 – ZTE Corporation Agree on these project properties.
6 – Futurewei We agree with the candidate project characteristics specified in Table 2 for Rel-18 SI on AI/ML for Air-Interface.
7 – InterDigital We support the proposal in Table 2
8 – Qualcomm Incorporated We support the proposal
9 – CAICT We support moderator's proposal.
10 – AT&T Support the proposal
11 – China Mobile Com. Corporation Support the proposal
12 – Intel Technology India Pvt Ltd Support
13 – CATT Support the proposal.

14 – Keysight Technologies UK Ltd We agree with the moderator's proposal.
15 – Guangdong OPPO Mobile Telecom. agree with the moderator's proposal
16 – MediaTek Inc. Support
17 – Apple Hungary Kft. Agree with the proposal
18 – Spreadtrum Communications Support FL's proposal
19 – Fujitsu Limited we support this proposal.
20 – Lenovo (Beijing) Ltd Agree on this proposal
21 – Xiaomi Communications We support this proposal
22 – LG Electronics Inc. Support
23 – FGI We support the proposal.
24 – ROBERT BOSCH GmbH Support
25 – NEC Corporation Agree
26 – Rakuten Mobile We agree with the proposal
27 – Deutsche Telekom AG Proposal is fine with us.

28 – Sony Group Corporation We are fine with the candidate project characteristics.
29 – Ericsson LM Ericsson supports the proposal.
30 – Nokia Corporation We support the proposal.
31 – HUAWEI TECHNOLOGIES Co. Ltd. Fine with it.

3.2 Candidate SID ”Justification”

Based on the discussion during the Initial round, I would like to suggest the following candidate SID justification:

Table 3: Justification section for candidate SID

The application of AI/ML to wireless communications has been thus far limited to implementation-based approaches, both, at the network and the UE sides. A study on enhancement for data collection for NR and ENDC (FS_NR_ENDC_data_collect) has examined the *functional framework for RAN intelligence enabled by further enhancement of data collection through use cases, examples etc. and identify the potential standardization impacts on current NG-RAN nodes and interfaces*. In this study, we explore the benefits of augmenting the air-interface with features enabling improved support of AI/ML based algorithms for enhanced performance. Enhanced performance here depends on the use cases under consideration and could be, e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead. A limited number of representative use cases will serve as a pilot to provide a broad view of applicability of AI/ML based techniques while enabling going deep into the evaluation and comparison versus traditional methods, as well as, characterizing the associated specification impact to enable those techniques. The goal is that the limited set of representative use cases will enable the identification of a common AI/ML framework which could be used for application to many other use cases in subsequent projects. The study will serve identifying what is required for an adequate AI/ML model characterization and description establishing pertinent notation for discussions and subsequent evaluations. Various levels of collaboration between the gNB and UE are characterized and considered. Evaluations to exercise the attainable gains of AI/ML based techniques for the use cases under consideration will be carried out with the corresponding identification of KPIs with the goal to have a better understanding of the attainable gains and associated complexity requirements. Finally, specification impact will be assessed in order to improve the overall understanding of what would be required to enable AI/ML techniques for the air-interface.

Please include your comments on the suggested ”Justification” section for the candidate SID:

Feedback Form 9: Inputs on ”Justification” section for candidate SID

1 – vivo Mobile Communication Co.

Thanks for the moderator’s comments.

The work in SA2 should also be mentioned for background information as following:

The application of AI/ML to wireless communications has been thus far limited to implementation-based approaches, both, at the network and the UE sides. A study on enhancement for data collection for NR and ENDC (FS*NR*ENDC*data*collect) has examined the functional framework for RAN intelligence enabled by further enhancement of data collection through use cases, examples etc. and identify the potential standardization impacts on current NG-RAN nodes and interfaces. **In SA WG2 AI/ML related study, a network functionality NWDAF (Network Data Analytics Function) was introduced in Rel-15 and was continuously enhanced in Rel-16 and Rel-17.**

2 – NTT DOCOMO INC.

We support the candidate SID justification.

Limiting representative use cases makes it possible to reduce workload and study representative use cases more in depth.

3 – SHARP Corporation

We are fine with the candidate SID justification.

4 – ZTE Corporation

We are generally fine with the above justification.

5 – VODAFONE Group Plc

We are fine with the candidate SID justification.

6 – Futurewei

In general, we are ok with the justification suggested by the moderator with the following minor feedbacks.

a) Typically, performance is associated with better accuracy, network/UE performance and/or reliability while reducing complexity and overhead usually come with a cost of performance, thus, they should be considered separately/differently. We suggest rephrasing the 2nd paragraph to something like “…AI/ML based algorithms for enhanced performance (e.g., improved throughput, robustness, accuracy or reliability) and/or reduced complexity/overhead”.

b) In Rel-18, only a limited set of use cases will be included, thus, it’s unlikely this SI will “provide a broad view of applicability of AI/ML based techniques”. However, we believe this SI will allow us to establish the basic/initial framework to support AI/ML-based approaches and operations for air interface, investigate common evaluation strategies and methodologies, and exercise the associated specification impacts via the identified use cases. This SI work will lay the foundation for future AI/ML Air Interface use cases while new AI/ML techniques will continue being explored and the framework will evolve together in supporting the new use cases and AI/ML techniques/operations.

Thus, we suggest rephrasing the sentence to something like "Through studying a few carefully selected use cases, assessing their performance in comparison with traditional methods and the associated specification impacts that enable their solutions, this SI will lay the foundation for future Air-Interface use cases leveraging AI/ML techniques."

7 – InterDigital

We are fine with the candidate SID justification.

8 – Qualcomm Incorporated

We are fine with the candidate SID justification.

9 – CAICT

We are generally fine with the proposed justification.

10 – AT&T

We are OK with the proposed justification.

11 – China Mobile Com. Corporation

For "A common AI/ML framework", does it mean a 3GPP framework to conduct study for AI/ML for air interface or a functional framework like the SI of FS*NR*END*C*data*collect applies? We think more clarification/description is needed.

12 – Intel Technology India Pvt Ltd

OK with text

13 – CATT

We are fine with the candidate justification.

14 – Keysight Technologies UK Ltd

We agree in principle with the proposed text by the moderator.

We would like to suggest adding "etc." to the following text in order not to exclude any other possible benefits from the application of AI/ML techniques: "e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead, etc.".

15 – Guangdong OPPO Mobile Telecom.

we are general fine with the justification

16 – MediaTek Inc.

We also find "a common framework" confusing, we understand this study is to start with representative use cases to understand the benefits and the architecture requirements of each use case. Please note that each use case may require a different AI/ML architecture solution e.g. using UE-based learning, or BS-based learning, offline learning etc. We can only discuss architecture after we conclude the study of each use case and also whether these requirements are applicable to other use cases. Since RAN2/3 typically use architecture, we suggest to use "architecture" instead of "framework" for avoid misunderstanding.

Wording suggestion

"...The goal is that the limited set of representative use cases will enable the identification of **functional requirements of AI/ML architecture** ~~which could be used for application to many other use cases in subsequent projects~~.

17 – Apple Hungary Kft.

We are fine with the proposed justification. May need some editorial updates.

18 – Spreadtrum Communications

Generally we are fine with FL's proposal.

19 – Fujitsu Limited

We are fine with the candidate SID justification.

20 – Lenovo (Beijing) Ltd

We are fine with the descriptions on the SID justification.

21 – ROBERT BOSCH GmbH

We are fine with the description

22 – FGI

We are fine with the candidate SID justification.

23 – NEC Corporation

Agree on the justification.

24 – Rakuten Mobile

We are fine for the candidate SID justification

25 – Sony Group Corporation

Thank you for drafting the justification.

We are basically fine with the justification. However, we think use cases shouldn't be down-selected at this stage. We should have an initial phase of the study to identify use cases, encouraging new use cases beside the afore-mentioned three use cases.

We would prefer that the justification emphasized the generally applicable benefits of the study rather than that there might be benefits of AI/ML to specific use cases. Hence, we think that the third paragraph should read:

The goal is that sufficient use cases will be considered to enable the identification of a common AI/ML framework which could be used in subsequent projects. The study should also identify areas where AI/ML could improve the performance of air interface functions.

We think the study should avoid "going deep into the evaluation and comparison versus traditional methods". Given the rapidly improving performance of AI/ML technology, we would expect comparisons with the current performance of functions to soon become out of date. Hence, comparison of performance versus traditional methods is not necessarily of prime importance in the study, rather 3GPP should be aiming

to make the system welcoming to the adoption of AI technologies. We should not be studying the use cases themselves, we should be studying what the use cases tell us about how AI/ML can be applied. For example, this is not a study about how to improve the CSI feedback function, it is a study that uses the CSI feedback example to allow us to work out the procedures that are required to replace or augment existing functions with AI/ML technology.

26 – Deutsche Telekom AG

The text for SI justification is fine with us.

27 – Ericsson LM

Ericsson supports the justification text.

28 – Nokia Corporation

We support the moderator's proposal for the justification.

29 – HUAWEI TECHNOLOGIES Co. Ltd.

1. Change "characterizing the associated specification impact to enable those techniques" in the third paragraph to "identifying the potential associated specification impact to enable those techniques". Firstly, whether there is any associated specification impact for a certain use case still depends on further study, therefore for safe we can add "potential" here. Secondly, for a SI, the most we can achieve is to identify the potential specification impact and then if needed continue to complete it in WI, therefore it seems "identifying" is more accurate than "characterizing", which may be more suitable for an WI.

2. Delete "The goal is that the limited set of representative use cases will enable the identification of a common AI/ML framework which could be used for application to many other use cases in subsequent projects". In our understanding, we already have the basic framework for the study of all those use cases, i.e. the one agreed under *FSNRENDCdatacollect* as described in the first paragraph, and also captured in section 4 in TR 21.905. The functional framework agreed under *FSNRENDCdatacollect* should be taken as the baseline, and whether any new modification needed can depend on the further study in the study phase. It seems different companies have different definition/understanding on the word "framework" also based on the discussion in the initial round, to avoid the confusing with what agreed in RAN3 SI, maybe we can use some other word like "procedure" to replace the "framework" in the SID, unless it exactly means the RAN3 agreed framework.

3. As described in point 2 above, we suggest to add a sentence in the justification describing that "For the study on AI/ML for air interface, basic framework agreed for *FSNRENDCdatacollect* as captured in section 4 in TR 21.905 is taken as baseline".

4. Delete "characterized and" in the sentence "Various levels of collaboration between the gNB and UE are characterized and considered". We think it is sufficient to use "are considered" at this stage. As to whether to character various levels, it would depend on the further study in the study phase.

3.3 Candidate SID "Objectives"

Based on the discussion during the Initial round, I would like to suggest the following candidate SID objectives:

Table 4: Objectives section of candidate SID

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance, complexity and potential specification impact.
Use cases to focus on:

- Initial set of use cases includes:
  - CSI feedback enhancement: overhead reduction, improved prediction with mobility
  - Beam management: beam prediction in time, spatial domain
  - Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH
- Finalize representative set of use cases for characterization and evaluations

AI/ML model and description:

- Characterize the defining stages of AI/ML related algorithms:
  - Model generation, e.g., initiation/configuration, training (online/offline), model transfer, as applicable
  - Inference
- Identify various levels of collaboration frameworks between UE and gNB, e.g.,
  - No collaboration: implementation-based only AI/ML algorithms
  - Various levels of UE/gNB collaboration targeting separate or joint ML operation.
- Lifecycle management of AI/Model: e.g., data collection, model training, model deployment, model monitoring, model updating

For the use cases under consideration:
1) Evaluate performance benefits of AI/ML based algorithms

- Methodology based on statistical models (from TR 38.901 and RE 38.857 [positioning]), link and system level simulations.
  - Extensions to 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
  - Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study.
  - Consider offline training vs. online training of models.
- KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels. Determine the use-case specific KPIs and benchmarks of the selected use-cases.
  - Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML) baseline
  - Overhead and memory storage associated with enabling respective AI/ML scheme should be well documented.

2) Assess specification impact specifically for the given use case and for a common framework:

- PHY layer specification impact including (RAN1)
  - Need for dataset construction for training, validation and test for the selected use cases
  - Use case and collaboration level specific specification impact

Please include your comments on the suggested ”Objectives” section for the candidate SID:

Feedback Form 10: Inputs on ”Objectives ” section for candidate SID

1 – vivo Mobile Communication Co.

Thanks for the moderators efforts. In general the SID objectives are in good formulation.

We would like to point out the following aspects:

1. Based on current formulation, our understanding is that there would be two phases, one phase with study on initial set and the second phase on the finalized set. The relationship between initial set and finalized set need to be clarified, including which of the following SID objectives is considered for initial set and which of the following objectives is considered for finalized set. Our understanding is as following

a) AI/ML model and description part is applicable for initial set

b) Evaluation and specification assessment part is applicable for finalized set

2. For the initial set:

a) We believe more cases can be included as pointed previously: channel estimation enhancement, mobility enhancement;

b) For the CSI use case, we would like to point out the overhead part should also include measurement RS overhead reduction;

c) For the beam management use case, we would also like to point out the possibility of reducing measured RS overhead for BM purposes.

3. For the finalized set evaluation, we think current formulation is good

4. For the finalized set specification impact, the description includes too detailed information that may not be needed at this stage.

5. For the finalized set specification impact, the impact assessed may not be limited to RAN1/RAN2, but may also be related to RAN3 or SA. Coordination with other TSGs/WGs may be necessary.

6. The intention of the statement “detailed AI/ML model descriptions is not expected to be specified and are left to implementation” need to be clarified. We believe the wording should be changed to “~~detailed/~~**specific** AI/ML model **used for each use case** ~~descriptions~~ is not expected to be specified and are left to implementation”

Based on the mentioned aspects, we have the following update of the SID **highlighted in bold:**

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance, complexity and potential specification impact.

Use cases to focus on:

1. Initial set of use cases includes:

a) CSI feedback enhancement: **payload and RS** overhead reduction, improved prediction with mobility

b) Beam management: **RS overhead reduction,** beam prediction in time, spatial domain

c) Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH

d) **Channel estimation enhancement: DMRS, CSI-RS, SRS**

2. Finalize representative set of use cases for characterization and evaluations

Clarify AI/ML model and description for initial set of use cases:

1. Characterize the defining stages of AI/ML related algorithms:

a) Model generation, e.g., initiation/configuration, training (online/offline), model transfer, as applicable

b) Inference

2. Identify various levels of collaboration frameworks between UE and gNB, e.g.,

a) No collaboration: implementation-based only AI/ML algorithms

b) Various levels of UE/gNB collaboration targeting separate or joint ML operation.

3. **Clarify** lifecycle management of AI/Model: e.g., data collection, model training, model deployment, model monitoring, model updating

For finalized representative set: ~~the use cases under consideration:~~

1. Evaluate performance benefits of AI/ML based algorithms

a) Methodology based on statistical models (from TR 38.901 and RE 38.857 [positioning]), link and system level simulations.

i) Extensions to 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.

ii) Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study.

iii) Consider offline training vs. online training of models.

b) KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels. Determine the use-case specific KPIs and benchmarks of the selected use-cases.

i) Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML) baseline

ii) Overhead and memory storage associated with enabling respective AI/ML scheme should be well documented.

2. Assess specification impact specifically for the given use case and for a common framework **considering lifecycle management of AI/Model, e.g., data collection, model training, model deployment, model monitoring, model updating (RAN1, RAN2)**

a) **Note: If needed, co-operate with other working groups for the assessment, e.g. RAN3/SA2.**

~~n PHY layer specification impact including (RAN1)~~

~~u Need for dataset construction for training, validation and test for the selected use cases~~

~~u Use case and collaboration level specific specification impact~~

~~n Protocol needs including (RAN2)~~

~~u Procedures related to training configuration~~

~~u Procedures related to interference configuration~~

~~u Procedures related to data management and model management~~

~~u AI/ML capability indication~~

3. Assess interoperability and testability aspects (RAN4)

Note: ~~**detailed/**~~**specific** AI/ML model**(s) used for each use case** ~~**descriptions**~~ is not expected to be specified and are left to implementation. User data privacy needs to be preserved.

2 – NTT DOCOMO INC.

We are generally fine with the candidate SID Objectives.

As one of expected gains brought by AI/ML beam management is overhead reduction with the small number of beam measurements and reports, "Beam management: beam prediction in time, spatial domain" can be modified into "Beam management: beam prediction in time and spatial domain, **overhead reduction**"

Also, in our views, further discussion in the study phase will determine whether or not field data should be used for performance assessments. It is safe to avoid stating "field data **should** be used" at this point.

3 – SHARP Corporation

We are fine with the proposal objectives in principle. In addition, we suggest the following corrections/-modifications in the text.

- Model generation, e.g., initiation/configuration, training (online/offline), model validation, model test, model transfer, as applicable
- ~~RE~~ TR38.857
- Procedures related to ~~interference~~ inference configuration

4 – MediaTek Inc.

RAN2 Chair: The work method described in the justification, first study a limited number of use cases and then attempt to identify the useful options, and then go ahead to identify common principles seems indeed to make sense as it gives some focus to the work. I haven't following this discussion in detail. I am worried that the RAN2 objectives will be interpreted very widely without focus, resulting in mostly nonproductive time consuming discussions.

1> Is it possible to rephrase to make RAN2 objectives a bit more focused? I think indeed it should be ok to apply the listed use cases also to the RAN2 work (I understand that this might be the intention).

2> If RAN2 shall initiate work on protocols an architecture before use cases has been analyzed, RAN2 will need to work based on blind architecture assumptions, which is not preferred but may be ok to some limited extent. Working on blind assumptions brings the risk that RAN2 spends majority of time working on options that are not useful to any of the listed use cases in the end, and it also brings the risk that RAN2 need to do double work with RAN1, E.g. where is learning done, where is inference done? In the UE, the gNB, the SMLC, other server?, where does the data come from? AI/ML is IMHO trial / error based - which entity decides to what extent a taken decision is the right one? is on-line learning practically feasible at all? IMHO all of these questions need to be addressed in a use-case specific way. Is there any way to organize the work, e.g. time-wise, e.g. clarify the work split between the groups etc, to reduce this obvious risk of double / wasted work?

5 – ZTE Corporation

We think this draft SID is a good baseline for further discussion. Some comments are given below.

Use cases:

We think it is sufficient to have these three use cases for Rel-18 study. Other use cases (e.g., RS overhead reduction) can be left for future releases. As these three use cases are to be studied carefully with use case characterization and evaluation, it seems not needed to have a further selection of representative set of use cases. Hence we can remove "Finalize representative set of use cases for characterization and evaluations", or reword it as "Finalize *the use cases under consideration* for characterization and evaluations".

We are supportive to have specific description to some extent. The current text generally looks good. For CSI feedback, we suggest to revise it as "*CSI compression feedback in spatial, frequency or time domain*" to align the understanding on the target of this CSI enhancement.

Evaluation:

"RE 38.857" should be "TR 38.857"?

We don't think field data is something "should" be used. Whether field data can verify the robustness of AI based approach is unclear, as to acquire a diverse set of field data to do this verification is nearly impossible. We suggest to revise this bullet as following.

Field data to further assess the performance and robustness in real-world environments *can be optionally used*.

Specification impact

We think the specification impact should involve RAN2, but it seems it can be done in future WI phase directly. The RAN2 specification impact may include necessary signaling and procedure enhancement to assist AI model life cycle management (e.g., model deployment/training/adaptation), but the detailed need and major functionality of such life cycle management should be studied in RAN1. Once RAN1 has conclusion on the need and functionality, it can be implemented in RAN2 during WI phase directly. We are not sure what needs to be studied in RAN2 before RAN1 concludes on the LCM. Hence we suggest to revise this RAN2 bullet as following, which can also avoid the risk of blind assumption or double work in RAN2 mentioned by RAN2 chair.

− PHY layer specification impact including (RAN1)

○ Need for dataset construction for training, validation and test for the selected use cases

○ Use case and collaboration level specific specification impact

Need and functionality of AI model life cycle management

~~- Protocol needs including (RAN2)~~

~~Procedures related to training configuration~~

~~Procedures related to interference configuration~~

~~Procedures related to data management and model management~~

~~AI/ML capability indication~~

6 – VODAFONE Group Plc

We are generally fine with the proposed objectives. As mentioned by Vivo, we also think the intent of the note "detailed AI/ML model descriptions is not expected to be specified and are left to implementation" should be clarified. It could also be argued that "preservation of user data privacy" is down to implementation and ultimately the responsibility of the vendor, so it isn't clear what we intend to achieve from a standards perspective by stating this without getting into implementation specifics.

7 – Futurewei

• As discussed in the first round, most companies agreed that studying a framework to support AI/ML for Air interface is one of the objectives for the "AI/ML for Air-Interface" SI, thus, we suggest adding this "common framework and principles for AI/ML for Air-Interface" to the objective section of the SID and including the following items:

a) Modifications or enhancements to the framework studied under Rel-17 RAN3 SI for "FS*NR*ENDC*data*collect" to support A/ML for air interface operations

b) Common principles in addition to those already specified under "FS*NR*ENDC*data*collect" and identify what changes are needed if any.

c) Common training/validation/testing assumptions and procedures for different collaboration levels.

d) Evaluation strategies/methodologies for complexity/overhead/performance, dataset construction/generation guidelines, baselines and other common aspects.

• Use cases to focus on: during round 1 discussion, many companies suggested supporting the original 5 use cases agreed from last email discussion as documented in [2], RP-211662, "Moderator's summary for discussion [RAN93e-R18Prep-12] AI/ML" (at least 10 companies suggested adding RS overhead reduction use case and 8 companies suggested adding mobility enhancement. Can the moderator kindly explain the reason why the other 2 use cases were removed from the list?

• For the use cases under consideration: in general, we agree the 2 aspects the moderator proposed to be studied for each use case, i.e., "evaluate performance benefits of AI/ML based algorithms" and "assess specification impact specifically for the given use case". However, some of the items indicated in the description belong to common framework or principles across use cases, e.g., "Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels". Thus, we suggest moving those "common" aspects to the "common framework and principles" section as we discussed in the first bullet of this topic.

At use case level, different feasible solution options should be considered and discussed, which may have different specification impacts, and even different performance benefits.

8 – InterDigital

We are generally OK with the proposed objectives. We assume that the CSI feedback enhancement includes also the RS overhead reduction in addition to feedback overhead reduction. We don't think the SID objective should mandate the usage of field data for evaluation.

9 – CAICT

Thanks for moderator's great efforts and the well-organized objectives.

We have the following comments on the proposed objectives:

1 use case to focus on

We propose "Channel estimation enhancement" could also be included in the initial set of use cases. Channel estimation is an important area for AI/ML models to be used and lots of studies have been published. Clear performance gain has also been shown in the 2nd WAIC with 3GPP channel model. AI/ML based channel estimation could be used not only for DMRS, but also PSS/SSS/CSI-RS/SRS. In general, AI/ML based channel estimation is an important direction for the development of AI/ML based air-interface enhancement.

2 AI/ML model and description

We are fine.

3 For the use cases under consideration:

1□Based on the objectives in use cases part, as vivo pointed out, this part should be clarified as "For the finalize representative set of use cases", since evaluation related works are use cases specific.

2□We proposed to have some revises on the field data part as:

Field data to further assess the performance and robustness in real-world environments should be ~~used. Associated details to be discussed as~~ part of the study.

3) As for the RAN2 related specification works, we also think it will be highly depended on the progress in RAN1. RAN2 related detail objectives could be defined after the initial set discussions in RAN1.

10 – AT&T

- We have concerns with the current use cases as described. As mentioned by other companies, RS overhead reduction and mobility enhancements should be included as standalone use cases, not combined or hidden in the other sub-bullets. Of course, there may be some commonality between CSI feedback enhancements and beam management and RS overhead reduction for example, but it will be much clearer to list them as separate use cases, as it can be seen that companies do not have a very clear understanding what is and what is not currently in scope. Given that there will be a consolidation and downselection phase as well, we do not see any risk with keep the original 5 use cases identified in the previous round of discussions.
- It is not clear exactly how field data can be used as part of evaluations in a straightforward manner in 3GPP. We have a concern that this is similar to the proposals to use traffic trace data in the Rel-17 XR SI, such that companies may expend significant effort to collect and prepare the data, which is ultimately never used because of the simulation complexity and difficulties in finding a common methodology baseline. We don't want to discourage the sharing or usage of real-world data, but it may be best considered as an "Identify how/whether" type objective at this stage.
- We strongly recommend to keep the RAN2 objectives. We support the comments of the RAN2 chair that early involvement of RAN2 in the study of use cases and the underlying common framework can be very beneficial if the work is well managed to avoid significant overlapping tasks with RAN1.

11 – Qualcomm Incorporated

We mostly agree with the candidate SID.

We suggest adding more clarification on this bullet:

PHY layer specification impact including (RAN1)

- Use case and collaboration level specific specification impact

by changing into

- Use case and collaboration level specific specification impact, such as new signaling, assistance information, measurement, and feedback

We agree with some companies' comment that the use of field data shouldn't be mandated at this stage.

We agree with some companies' comment on explicitly mentioning overhead reduction for CSF and beam management.

12 – China Mobile Com. Corporation

Use cases:

Although there are several interesting use cases, we should manage a reasonable scope of the SI. Based on the comments, there are different categories and different understandings for each use case. For example, CSI feedback enhancement (estimation and prediction) includes CSI feedback compression and sparse CSI feedback. For positioning, AI/ML can be applied to identify LOS/NLOS or estimate UE position directly. More use cases and categories means more efforts on performance evaluation, we think at most 2 use cases should be selected to conduct performance evaluation. Therefore, we suggest the following update:

− Finalize **1 2** representative set of use cases for characterization and evaluations

Field date:

Now the feasibility of using field data is questionable. Suggest the following update:

-Whether field data to further assess the performance and robustness in real-world environments should be discussed as part of the study.

Regarding the Note, we agree that detailed AI/ML model descriptions is not expected to be specified and are left to implementation. But we think at study phase, a typical or reference model can be provided for calibration and seen as a insight for AI/ML model design. Suggest the following update:

Note: detailed AI/ML model descriptions is not expected to be specified and are left to implementation. **Typical AI model(s) can be provided for calibration.** User data privacy needs to be preserved.

13 – Intel Technology India Pvt Ltd

The second bullet for use-cases is not needed, "Initial set" already means that WGs have certain flexibility in terms of use-case. we propose some wording change for CSI and beam-management. First bullet of "AI/ML model and description" is not needed – last bullet already captures it. Offline/online training discussion can be left to RAN1 to discuss. Propose common data set construction and the related bullet can be moved from spec impact to evaluation. Pls. see proposal below:

Use cases to focus on:

- Initial set of use cases includes:
  - CSI feedback enhancement: overhead reduction, precoder determination in spatial, frequency, time domain ~~improved prediction with mobility~~
  - Beam management: beam prediction ~~in time, spatial domain~~ with reduced measurement/reporting
  - Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH

  ~~Finalize representative set of use cases for characterization and evaluations~~

AI/ML model and description:

~~- Characterize the defining stages of AI/ML related algorithms:~~

  - ~~Model generation, e.g., initiation/configuration, training (online/offline), model transfer, as applicable~~
  - ~~Inference~~

- Identify various levels of collaboration frameworks between UE and gNB, e.g.,
  - No collaboration: implementation-based only AI/ML algorithms
  - Various levels of UE/gNB collaboration targeting separate or joint ML operation.
    - Lifecycle management of AI/Model: e.g., data collection, model training, model deployment, model monitoring, model updating

    1) Evaluate performance benefits of AI/ML based algorithms

    Methodology based on statistical models (from TR 38.901 and RE 38.857 [positioning]), link and system level simulations.
  - Extensions to 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
  - Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study.
  - Need for common dataset construction for training, validation and test for the selected use cases

- ~~Consider offline training vs. online training of models.~~

  KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels. Determine the use-case specific KPIs and benchmarks of the selected use-cases.
- Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML) baseline
- Overhead and memory storage associated with enabling respective AI/ML scheme should be considered ~~well documented.~~

  2) Assess specification impact specifically for the given use case and for a common framework:

- PHY layer specification impact including (RAN1)

- ~~Need for dataset construction for training, validation and test for the selected use cases~~
- Use case and collaboration level specific specification impact

---

14 – Keysight Technologies UK Ltd

We support the general principles stated in the SID objectives proposed by the moderator. We do have some comments and proposed modifications.

Regarding the **"Initial set of use cases"** section:

- It is not clear what the objective of "Finalize representative set of use cases for characterization and evaluations" is. It should be clarified what aspects of the use cases need to be finalized as part of the SI. Is the intention to leave the SI open for additional use cases? We don't think we should open the door for many more different use cases just to maintain scope under control. In any case, whether this list of use cases is considered closed or open, should be explicitly mentioned.

Regarding the **"AI/ML model and description"** section:

- We would like to suggest adding model validation and model testing to these 2 objectives:
  - "Model generation, e.g., initiation/configuration, training (online/offline), model validation, model testing, model transfer, as applicable"
  - "Lifecycle management of AI/Model: e.g., data collection, model training, model validation, model testing, model deployment, model monitoring, model updating"
- It is not clear what the SI objective will be for the "No collaboration" framework. What is the specific objective for the study item under this no-collaboration framework? Will it involve air interface changes not related to collaboration (i.e., training and inference)? As mentioned previously, it might be needed to clarify that the meaning of "no collaboration" is "signaling not involved in training and inference although any other changes in the air interface might be needed depending on the outcome of the study item".

Regarding the **"For the use cases under consideration"** section:

- Keysight would like to reword "for the given use case" to "for the agreed use cases…"
- Regarding "Protocol needs including (RAN2)":
  - We would like to propose removing "configuration" from the first two bullets. At this stage, we don't think we should limit the protocol impact for training and inference to just configuration as there might be others.
  - We would like to add "Procedures related to training, validation and testing".

Regarding the last **"Note"**:

- We would like to suggest moving "User data privacy needs to be preserved" to the "For the use cases under consideration" section as: "Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study. User data privacy needs to be preserved.".
  - We think we should highlight that this aspect will need to be discussed and provide recommendations during the SI. Otherwise, this important aspect required for operationalizing AI/ML in networks may not be handled.

15 – Guangdong OPPO Mobile Telecom.

As for use case CSI feedback enhancement, we are not sure what does "improved prediction with mobility" really mean. Since it is only mentioned by very few company, we propose to remove the wording. As for beam management, we still believe RS reduction is better use case. but we can also accept not to cover both as compromise.

we are still not convinced that data set from field is necessary at this stage and hence propose to remove it now from objectives.

As for spec impact, apart from PHY layer, we should also include RAN2, especially on AI/ML model management aspects.

here are the revised objectives from our side with some more editorial change:

.....

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects

such as performance, complexity and potential specification impact.

Use cases to focus on:

− Initial set of use cases includes:

- CSI feedback enhancement: overhead reduction~~, improved prediction with mobility~~
- ~~Beam management: beam prediction in time, spatial domain~~
- Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH

− Finalize representative set of use cases for characterization and evaluations

AI/ML model and description:

− Characterize the defining stages of AI/ML related algorithms:

- Model generation, e.g., initiation/configuration, training (online/offline), ~~model transfer,~~ as applicable
- Inference

− Identify various levels of collaboration frameworks between UE and gNB, e.g.,

- ~~No collaboration: implementation-based only AI/ML algorithms~~
- Various levels of UE/gNB collaboration targeting separate or joint ML operation.

− Lifecycle management of AI/Model: e.g., data collection, ~~model training,~~ model deployment, model monitoring, model updating, **model transfer**

For the use cases under consideration:

1) Evaluate performance benefits of AI/ML based algorithms

− Methodology based on statistical models (from TR 38.901 and RE 38.857 [positioning]), link and system level simulations.

- Extensions ~~to~~ **of** 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
- ~~Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study.~~
- Consider offline training vs. online training of models.

− KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels. Determine the use-case specific KPIs and benchmarks of the selected use-cases.

- Performance and complexity of AI/ML based algorithms should be compared to that of a stateof-the-art (non-AI/ML) baseline
- Overhead and memory storage associated with enabling respective AI/ML scheme should be well documented.

2) Assess specification impact specifically for the given use case and for a common framework:

− ~~PHY layer~~ specification impact including (RAN1/**RAN2**)

- ~~Need for dataset construction for training, validation and test for the selected use cases~~
- Use case and collaboration level specific specification impact

16 – MediaTek Inc.

We suggest following revisions.

On use case

- CSI feedback enhancement: overhead reduction, improved prediction with mobility **(RAN1)**
- Beam management: beam prediction in time, spatial domain (RAN1)
- Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH **(RAN1)**
- **RRM mobility: prediction, overheard reduction (RAN2)**
- ~~Finalize representative set of use cases for characterization and evaluations~~

On spec impact

2. Assess specification impact specifically for the given use case **considering lifecycle management of AI/Model, e.g., data collection, model training, model deployment (RAN1, RAN2)**

17 – Apple Hungary Kft.

o For use case, we support the RS overhead reduction with AI based channel estimation over the positioning use case. And we agree with other companies reference signal overhead reduction should be included for CSI use case and beam management use case.

o For AI/ML model description, propose to add one bullet under "Characterize the defining stages of AI/ML related algorithms"

- *Data set for training/validation: including raw data generation based statistical models or field data, and data preprocessing*

o "Lifecycle management of AI/Model: e.g., data collection, model training, model deployment, model monitoring, model updating": Since AI/ML has been used in various use cases spanning from core network to RAN3, RAN2 and RAN1, lifecycle management of AI models can be generic and potentially the part of O&M. It is not clear the proposal here is RAN1 use cases have its lifecycle management of AI models, while other RAN/CN will have different life cycle management.

o On specification impact, we believe RAN2 scope needs to be clarified, as AI lifecycle management can evolve too many aspects as commented above.

o PHY layer specification impact including (RAN1)

§ ~~Need for dataset construction for training, validation and test for the selected use cases~~

§ ~~Use case and~~ collaboration level specific specification impact **per use case**

o **Signaling design to support the collaboration identified in RAN1. (RAN2)**

o Interoperability and testability aspects (RAN4)

18 – CATT

Thanks for the summary. Please find CATT comments below.

Use case:

We are ok to include the three use cases in the initial set. But we prefer to limit the number of use case for evaluations, as work load increases linearly with the number of use cases for detailed study. We propose the following revision:

Finalize representative set of 1 2 use cases for characterization and evaluations

AI/ML model and description:

Online training involves complicated message exchanging between network and UE and shall be avoided in this release. After we have better understanding on AI/ML application based on offline training, online training can be considered in future releases. The following revision is suggested:

Model generation, e.g., initiation/configuration, training (~~online/~~offline), model transfer, as applicable

~~Consider offline training vs. online training of models.~~

KPI:

Implementation-based AI/ML algorithms shall also be considered as baseline for performance characterization.

Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML and implementation-based AI/ML) baseline

Specification impact:

In our understanding, each of the finalized representative use case belongs to one specific category and collaboration level. Studying use case specific specification impact is sufficient. The following revision is suggested:

Use case ~~and collaboration level~~ specific specification impact

Regarding RAN2 aspects, it seems detailed study on procedure/signaling can be conducted only after RAN1 has identified performance gain for the respective use case. This can be done in follow-up WI.

19 – Spreadtrum Communications

In principle we are fine with moderator's proposal. Given there are many issues for field data, e.g., how to collect it, how to identify the validation it, in our view, it should be optional, not mandatory as one tool to assess the performance. Regarding the standardization work for RAN2, we think indeed it should be involved. But it is reasonable to kick off RAN2 work, only after the clear picture has been achieved in RAN1, to avoid double work mentioned by RAN2 Chair. Maybe the specific standardization impact of RAN2 can be captured in the SID in later stage.

20 – Fujitsu Limited

Use cases,

We also suggest to add channel estimation (targeting to reduce RS overhead) as one of Initial set of use cases;

Field data,

we don't think it is already concluded field data 'should be used' to further assess the performance and robustness. The methodology and feasibility to use filed data in training/testing AL models should be studied first if necessary.

RAN2 work,

We also think it is depended on the results/progress of RAN1 initial study. RAN2 relevant detials is suggested to be removed.

21 – Lenovo (Beijing) Ltd

In general, we are fine on the objectives. Because "the goal is … the identification of a common AI/ML framework…" as described in the Justification, the investigation on a framework (or architecture) over the 3GPP air-interface to support various AI/ML types (i.e., different requirements of collaboration levels) could be more meaningful than doing evaluations on more use-cases, if some use-cases may use the same kind of AI/ML type, e.g., auto-encoder. Thus, we think the representative set of use-cases make sense if they can cover as much typical AI/ML types as possible. In addition, We don't think it is "should" to use the field data, because there will be additional assessment work on the 'field data' (e.g., good or not), which may have risk to converge. Thus, we suggest not to consider 'field data' within the evaluation methodology or at least to discuss whether it is really needed and workable.

We further have following comments/updates on the objectives:

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects, such as performance, complexity and potential specification impact.

Use cases to focus on:

- Initial set of use cases includes:
  - CSI feedback enhancement, **e.g.**, overhead reduction, improved prediction~~ with mobility~~
  - Beam management, **e.g.,** beam prediction in time **and/or** spatial domain**, overhead and latency reduction**
  - Positioning accuracy enhancements, **e.g.,** in scenarios with high NLOS probability~~, e.g., IIOT DH~~
- Finalize representative set of use cases for characterization and **baseline performance** evaluations

AI/ML model and description:

- Characterize the defining stages of AI/ML related algorithms:
  - Model generation, e.g., initiation/configuration, training (online/offline), model transfer, as applicable
  - Inference **operation, e.g., input/output, pre-/post-process, as applicable**
- Identify various levels of collaboration ~~frameworks~~ between UE and gNB, e.g.,
  - No collaboration: implementation-based only AI/ML algorithms
  - Various levels of UE/gNB collaboration targeting separate or joint ML operation.
- Lifecycle management of AI/M**L m**odel: e.g., data collection, model training, model deployment, model monitoring, model updating

For the use cases under consideration:

1) Evaluate performance benefits of AI/ML based algorithms

- Methodology based on statistical models (from TR 38.901 and ~~RE~~ **TR** 38.857 [positioning]) **for** link and system level simulations.
  - Extensions to 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
  - **Whether** field data **are needed** to further assess the performance and robustness in real-world environments should be ~~used~~ **discussed**. ~~Associated details to be discussed as part of the study.~~
  - ~~Consider offline training vs. online training of models.~~
- KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with different collaboration levels. Determine the use-case specific KPIs and benchmarks~~ of the selected use-cases~~.
  - Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML) baseline
  - Overhead and memory storage associated with enabling respective AI/ML scheme should be well documented.

2) Assess specification impact specifically for the given use case and for a common framework:

- PHY layer specification impact including (RAN1)
  - Need for dataset construction for training, validation and test ~~for the selected use cases~~
  - ~~Use case and~~ Collaboration level specific specification impact**, such as new reference signals**

- **Upper layers specification impact** ~~Protocol needs~~ including (RAN2)
  - **New procedures to control and manage the AI/ML operations, e.g., to initiate and configure the model and inference**
  - ~~Procedures related to training configuration~~
  - ~~Procedures related to interference configuration~~
  - ~~Procedures related to data management and model management~~
  - AI/ML capability indication
- Interoperability and testability aspects (RAN4)

Note: detailed AI/ML model descriptions is not expected to be specified and are left to implementation, **and user data privacy should be preserved for any AI/ML operation.** ~~User data privacy needs to be preserved.~~

22 – Xiaomi Communications

1. For the use case , We are OK to focus on limited number of use cases. But there is no concret critera to judge which use case is ”representive” or ”not representive” . We think at this stage the channel estimation and the mobility should be included and we can narrow down the use case based on more concrete critera.

2. For the KPIs, in our view, the inference on the UE side would also cause impact on the device power consumption. Thus, the power consumption KPI should be included as well

23 – Samsung Electronics Co.

We have questions for 1st bullet, at the initial stage, many companies suggested RS overhead reduction and other use cases, e.g., non-linear handling, so that we suggest not to preclude there aspect in the summary. With the bullet of “Finalize representative set of use cases for characterization and evaluations”, does it mean other use cases are not precluded, if so we suggest making it more clear.

Besides, we suggest to further clarify some use cases, as below:

- Initial set of use cases includes:

o CSI feedback enhancement: overhead reduction, **reconstructed accuracy**, improved prediction with mobility

o Beam management: beam prediction in time, spatial domain **for UE RX beam tracking and gNB TX beam selection, joint beam pair recovery**

o Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH

o **Other use cases are not precluded**

More clarification on AI/ML model description and typo correction is suggested as below:

AI/ML model and description:

- Characterize the defining stages of AI/ML related algorithms:

o Model generation, e.g., initiation/configuration, training (online/offline), model transfer, as applicable

o Inference**, e.g. model deployment and activation, AI algorithm execution**

- Identify various levels of collaboration frameworks between UE and gNB, e.g.,

o No collaboration: implementation-based only AI/ML algorithms **without information interaction**

o Various levels of UE/gNB collaboration targeting separate or joint ML operations.

- Lifecycle management of AI/ML Model: e.g., data collection, model training, model deployment, model monitoring, model updating

Regarding "Field data to further assess the performance and robustness in real-world environments **should** be used. Associated details to be discussed as part of the study." We think 'should' is too strong wording. Depending on use cases, the evaluation with field data may not be so critical. Thus, 'could' could be a better wording. I think at a high level, no one disagrees with the use of field data. The more important question is 'how'. In this regard, we can make one small step further by revising, "Associated details, such as types of the field data, potential provider of the data, and how to host the data, to be discussed as part of the study."

Moreover, to better organize the study in the working group, we suggest to divide the study into two-phase. Phase I focuses on 1) evaluation performance, and make "2) assess specification impact…" as a second study phase. Without clear use cases, and study outcomes, it is hard for RAN 1/2/4 to assess specification impact.

For specification impact, we would like to know how to interpret the objective for RAN4 part:

- What kind of "Interoperability" is intended here for RAN4 to study?

- For "testability", usually in RAN4 terminology, "testability" usually refers to study the aspects of test equipment, procedure, methodology, measurement uncertainty, and other issues related to conformance testing. Need further clarification on what testability is expected from RAN4 perspective for AI/ML air interface.

24 – LG Electronics Inc.

Regarding the use case, we also prefer to include mobility enhancement. Similarly to some other companies, we have question/comment on the suggested work definition of each WG. Since this is not for WI but for SI, we wonder why procedures related to training/inference/model management should be handled by RAN2 only and why use case and collaboration level specific specification impact should be handled by RAN1 only. With the proposed work split, it would be difficult to manage interaction between RAN1 and RAN2 because performance and specification impact for each use case could depend on AI training/inference/model management procedure. In addition, we suggest RAN2 to discuss RAN2-centric use case(s) and related specification impact(s) as well, e.g. for mobility enhancement.

25 – MediaTek Inc.

On R4 scope, it is not clear to us what R4 can do with "Interoperability and testability aspects" evaluation before there is conclusion on use case study. Such task can be postponed to later WI phase.

26 – ROBERT BOSCH GmbH

In general, we think the objectives are already in a nice shape. We agree with LG that mobility enhancements is an additional interesting use case that should be included.

27 – NEC Corporation

For "Initial set of use cases", we want to clarify what the "overhead reduction" includes. As pointed out in the initial round, RS overhead reduction has been mentioned by many companies. So, we suggest to include it in the initial study

28 – FGI

- Identify various levels of collaboration frameworks between UE and gNB, e.g.,

  - No collaboration: implementation-based only AI/ML algorithms

  - Various levels of UE/gNB collaboration targeting **at** separate or joint ML operation.

- Lifecycle management of **AI/ML Model** ~~AI/Model~~: e.g., data collection, model training, model deployment, model monitoring, **and** model updating

29 – Rakuten Mobile

We are generally fine with the candidate SID objectives except for the initial set of use cases. For the initial set of use cases, possible candidates (at least channel estimation and mobility enhancement) should be on the table, and then we can downselect / merging the use cases.

30 – Sony Group Corporation

Thank you for drafting the objectives.

As we also mentioned in initial round and justification part, we should have initial phase of the study to identify use cases, which should not be limited to or downscoped from the initial set of use cases but should be broader. The initial phase of the study identifies potential use cases with a standards impact.

We are fine with other parts.

For the objectives on "evaluate performance benefits of AI/ML based algorithms", an important aspect is that we study the evaluation methodology and KPIs themselves, as opposed to focusing on the performance of the use cases themselves. The text in this section is OK as it is, but the focus of the evaluation study should not be on working out "how many dB better an AI/ML algorithm is to a traditional algorithm", it should be on how AI/ML algorithms in general are evaluated.

31 – Deutsche Telekom AG

The description of the SID objectives is generally fine with us, but we would prefer not to down-select at this stage the number of use cases to only three. Keeping the other 2 use cases under discussion (channel estimation/mobility) would give a broader view on needed AI/ML functions and model handling. Later down-selection should not be excluded, if required.

32 – Ericsson LM

We would like to clarify *"Initial set of use cases includes"*

- Does this imply that new use cases can be added within the Rel-18 study item?
- The scope of the SI is already large, involving many new issues for 3GPP. To ensure the greatest chance for success, we support limiting the number of use cases. Of course, this does not preclude other use cases in subsequent SI/WIs.

We would like to clarify *"Finalize representative set of use cases for characterization and evaluations"*.

- Does this imply that the SI will start with an initial use case selection phase?

Change *"Positioning accuracy enhancements in scenarios with high NLOS probability, e.g., IIOT DH"* to *"Positioning accuracy enhancements for different scenarios, including scenarios with high NLOS probability, e.g., IIOT DH,"* LOS scenarios are not precluded.

Change *"Field data to further assess the performance and robustness in real-world environments should be used. Associated details to be discussed as part of the study"* to *"Field data to further assess the performance and robustness in real-world environments can be used. Associated details to be discussed as part of the study."*

- Note: It is already possible to share results and conclusions drawn from field data, so this statement is not strictly needed in the SID.
- The SID should not force use of field data. Justifications for using field data, for given use cases, can be made by companies during the SI including procedures on if and how they intend to share their field data to 3GPP.

33 – Nokia Corporation

The moderator's proposal for the objectives looks quite ok but there are few areas, which in our view require updates.

As we are discussing initial set of use cases, which we aim to use for defining 3GPP framework for AI/ML, we need to conduct further technical studies before we can define details for the initial set of use cases and also the final set of use cases for the study. Thus, in our view only high level definition for the initial set of use cases can be defined at this point for the objectives. Also, companies are likely to have rather different understanding of the proposed use case details, which will complicate working group discussions even further.

In our view also the RAN4 objectives need to be more detailed like for RAN1 and RAN2. It is important for RAN4 to study UE requirement and testing framework both to validate AI/ML based performance enhancements and to ensure that UE meets the existing minimum requirements with AI/ML.

Thus, we propose the following updates to the objectives:

Use cases to focus on:

- Initial set of use cases includes:
  - CSI feedback enhancement~~: overhead reduction, improved prediction with mobility~~
  - Beam management~~: beam prediction in time, spatial domain~~
  - Positioning accuracy enhancements in scenarios with high NLOS probability~~, e.g., IIOT DH~~
- Finalize representative set of use cases for characterization and evaluations

2) Assess specification impact specifically for the given use case and for a common framework:

- PHY layer specification impact including (RAN1)
  - Need for dataset construction for training, validation and test for the selected use cases
  - Use case and collaboration level specific specification impact
- Protocol needs including (RAN2)

- Procedures related to training configuration
- Procedures related to interference configuration
- Procedures related to data management and model management
- AI/ML capability indication

- Interoperability and testability aspects (RAN4)
  - UE requirement and testing framework to validate AI/ML based performance enhancements and ensuring that UE with AI/ML meets the existing minimum requirements

34 – HUAWEI TECHNOLOGIES Co. Ltd.

1. Add "Support of AL/ML for air-interface shall reuse the existing RAN architecture and new interfaces shall not be introduced" to the note, i.e. update the note to "Note: detailed AI/ML model descriptions is not expected to be specified and are left to implementation. User data privacy needs to be preserved. **Support of AL/ML for air-interface shall reuse the existing RAN architecture and new interfaces shall not be introduced**". As we commented before, the principles identified during the SI of FS*NR*END*Cdata*collect should be applied by default, in order to do smooth interaction with the features developed for AI/ML for higher layer use cases.

2. For objective 2):

- Change "Assess specification impact specifically for the given use case and for a common framework" to "Assess **potential** specification impact **and aspects only for implementation,** specifically for the given use case and for a common **procedure**".
- We also prefer to delete the sub-bullet "PHY layer specification impact including (RAN1)" and the sub-bullet "Protocol needs including (RAN2)rotocol needs including (RAN2)", and then make the bullet for RAN4 as a separate objective.

Firstly, whether/what specification impact needed still depend on further study, so we need to add "potential" here. Secondly, we think it is also very necessary to make it clear which aspects are only for implementation and which aspects are for normative work, and thus "aspects only for implementation" is added. Thirdly, changing "framework" to "procedure" to avoid the confusion with the basic framework agreed in RAN3 SI. Finally, all the details can be discussed and identified during the study phase.

3. Change "framework" to "procedure" in the sentence "Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance". As described above, if we talk about framework, then the functional framework agreed under *FSNRENDCdatacollect* as captured in section 4 in TR 21.905 is already there for utilization, no need to specifically study on it. Of course during the study of use cases, if any modification needed then it can be further considered, but no need to set up a specification objective for it. However, based on the structure of the objectives, it seems the meaning of the "framework" here mainly means the following aspects, e.g. bullets under "Use cases to focus on:", "AI/ML model and description:" and "For the use cases under consideration:", if it is true then better to change the word "framework" to something else like "procedure", to avoid the confusion with what agreed in RAN3 SI.

4. For initial set of use cases includes:

a) CSI feedback enhancements: 1) delete "improved prediction with mobility" since it seems not belong to CSI feedback enhancement, e.g. gNB can do the predication also thus we don't think it belongs to CSI feedback enhancements; 2) add "improved accuracy". As we commented before that the CSI feedback study should not only focus on overhead reduction.

b) Delete "with high NLOS probability", since IIOT scenario with high NLOS probability is not the typical case, which can be seen also from the scenario for Rel-17 positioning.

c) Add “Channel estimation and RS enhancement: DMRS, etc” as another use case. Based on the discussion in [RAN93e_R18Prep-08], channel estimation and RS enhancement achieved broad interest also, just similar as the 3 use cases listed here, therefore no reason to only preclude RS enhancements. In addition, it is expected that the potential gain achieved for RS enhancements can be similar as other cases like CSI feedback enhancements also.

5. “Finalize representative set of use cases for characterization and evaluations” should be deleted, since evaluations should be done before finalizing representative set of use cases, in order to specify AI for really beneficial use cases.

6. Change “frameworks” to “modes” in the sentence “identify various levels of collaboration frameworks between UE and gNB”, to avoid the confusion with the basic framework agreed in RAN3 SI. We are fine to remove all the following detailed sub-bullets also, since that should depend on further study in the study phase.

8. For “Overhead and memory storage associated with enabling respective AI/ML scheme should be well documented”, suggest to add “power consumption” as one of the aspects to consider also since as common understanding power consumption is one critical factor also. Therefore, suggest to modify it to “Overhead, power consumption and memory storage associated with enabling respective AI/ML scheme should be well documented”.

4 Final Round

The Moderator would like to thank all the participants for their inputs.

We have unanimous support for the candidate project characteristics as copied below:

Table 5: Candidate project properties

Title	SI/WI	Primary WG	Secondary WG	Notes
SI on AI/ML for Air-Interface	SI	RAN1	RAN2, RAN4	SI for entire Rel-18 duration.

As a result, the Moderator will use them for the pre-allocated Tdoc# for this SID in RP-212708.

For the rest of NWM discussions, the focus will be in fine tuning the Justification and Objectives sections of this candidate project.

4.1 Candidate SID ”Justification”

Based on the many inputs received and without getting into one-by-one replies, the Moderator would like to propose the following revised text for further inputs during this final round of discussions. Unfortunately, NWM does not support revision marks. Therefore, the Moderator is pasting here the clean version of the revised text and the revision marks with changes from the previous version can be seen at:

www.3gpp.org / ftp / tsg_ran / TSG_RAN / TSGR_94e / Inbox / Drafts / [RAN94e-R18Prep-08]

Table 6: Justification section of candidate SID

The application of AI/ML to wireless communications has been thus far limited to implementation-based approaches, both, at the network and the UE sides. A study on enhancement for data collection for NR and ENDC (*FS_NR_ENDC_data_collect*) has examined the *functional framework for RAN intelligence enabled by further enhancement of data collection through use cases, examples etc. and identify the potential standardization impacts on current NG-RAN nodes and interfaces*. In SA WG2 AI/ML related study, a network functionality NWDAF (Network Data Analytics Function) was introduced in Rel-15 and has been enhanced in Rel-16 and Rel-17.
In this study, we explore the benefits of augmenting the air-interface with features enabling improved support of AI/ML based algorithms for enhanced performance (e.g., improved throughput, robustness, accuracy or reliability) and/or reduced complexity/overhead. Enhanced performance here depends on the use cases under consideration and could be, e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead, etc.
Through studying a few carefully selected use cases, assessing their performance in comparison with traditional methods and the associated potential specification impacts that enable their solutions, this SI will lay the foundation for future Air-Interface use cases leveraging AI/ML techniques.
The goal is that sufficient use cases will be considered to enable the identification of a common AI/ML framework, including functional requirements of AI/ML architecture, which could be used in subsequent projects. The study should also identify areas where AI/ML could improve the performance of air interface functions.
The study will serve identifying what is required for an adequate AI/ML model characterization and description establishing pertinent notation for discussions and subsequent evaluations. Various levels of collaboration between the gNB and UE are identified and considered.
Evaluations to exercise the attainable gains of AI/ML based techniques for the use cases under consideration will be carried out with the corresponding identification of KPIs with the goal to have a better understanding of the attainable gains and associated complexity requirements.
Finally, specification impact will be assessed in order to improve the overall understanding of what would be required to enable AI/ML techniques for the air-interface.
For the study on AI/ML for air interface, the basic framework agreed for *FS_NR_ENDC_data_collect*, as captured in section 4 of TR 37.817, should be taken into consideration for possible applicability.

While the Moderator appreciates the comment suggesting to take the general framework developed as part of the RAN3-led SI on *FS_NR_ENDC_data_collect* as the baseline for this project, he finds it premature to do so. Instead, the proposal is to take it into consideration for its applicability. It is worth remembering that the *FS_NR_ENDC_data_collect* project is meant for inter-network-node protocols without air-interface involvement, and this project is looking into AI/ML applied to the air-interface which spans a completely different scope. Certainly, there could be some principles fully applicable but going over a blanket adoption of the developed principles for a different scope without proper review seems unfounded.

Please include your comments on the suggested revised ”Justification” section for the candidate SID:

Feedback Form 11: Inputs on ”Justification” section for candidate SID

1 – vivo Mobile Communication Co.

Support.

2 – NTT DOCOMO INC.

Support the updated Justification.

3 – VODAFONE Group Plc

We support the updated justification.

4 – Keysight Technologies UK Ltd

We agree with the proposed text by the moderator.

We would like to suggest the following modifications to the proposed text:

- In this study, we explore the benefits of augmenting the air-interface with features enabling improved support of AI/ML based algorithms for enhanced performance ~~**(e.g., improved throughput, robustness, accuracy or reliability)**~~ and/or reduced complexity/overhead. Enhanced performance here depends on the use cases under consideration and could be, e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead, etc.

We agree with the moderator regarding the framework coming out of the RA3-led SI. In this sense, we would like to suggest the following text modification:

- For the study on AI/ML for air interface, the basic framework **and principles** agreed for FS_NR_EN_DC_data_collec as captured in section 4 of TR 37.817, should be taken into consideration for possible applicability.

5 – ZTE Corporation

We are fine with the updated justification.

6 – Charter Communications

We are fine with the updated justification.

7 – AT&T

We are fine with the proposed justification section.

8 – Deutsche Telekom AG

Propose justification text is fine with us.

9 – ROBERT BOSCH GmbH

Support

10 – Futurewei

In general, we are ok with the proposed text for the "Justification" except one minor comment:

- In the second paragraph, the last sentence "Enhanced performance here depends on the use cases under consideration and could be, e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead, etc." is redundant as these examples have been mentioned in the previous sentence; thus, this sentence can be removed.

11 – CAICT

We are fine with the proposed justification.

12 – InterDigital

We support the updated justification.

13 – Qualcomm Incorporated

We support.

14 – Apple Hungary Kft.

We support

15 – China Mobile Com. Corporation

We support

16 – LG Electronics Inc.

We are fine with the proposed justification

17 – Lenovo (Beijing) Ltd

We are fine with the updated justification.

18 – Verizon UK Ltd

We are fine with the proposed justification.

19 – Xiaomi Communications

We support the updated justification

20 – Fujitsu Limited

We support.

21 – Rakuten Mobile

We support the Justification

22 – Ericsson LM

We support.

23 – Sony Group Corporation

Thank you for updating the Justification and considering our comments.

We are basically fine with the Justification.

Given the rapidly improving performance of AI/ML technology, we would expect comparisons with the current performance of functions to soon become out of date. Hence, comparison of performance versus traditional methods is not necessarily of prime importance in the study, rather 3GPP should be aiming to make the system welcoming to the adoption of AI technologies. So, we might not need to say "… in comparison with traditional methods" in the Justification (i.e., assessing their performance ~~in comparison with traditional methods~~, ..).

We are also supportive to the moderator's thought that this SI should not be bounded to the current findings in the *FS_NR_ENDC_data_collect* project, due to the difference in the scopes of studies.

24 – Spreadtrum Communications

We support

25 – Nokia Corporation

We support the moderator's justification proposal.

26 – MediaTek Inc.

We support.

27 – Continental Automotive GmbH

We support the updated justification.

28 – FGI

We support the updated Justification.

29 – HUAWEI TECHNOLOGIES Co. Ltd.

We appreciate/respect the explanation from the moderator, but we don't agree with it. It is true that the RAN3 AI/ML study mainly focus on higher layer use case while the RAN1 AI/ML study mainly focus on physical layer use case, however this should not be used to justify different framework for air-interface use case and higher layer use case in Rel-18. Same framework for all AI/ML use cases can ensure smoothly interaction whenever needed. In addition, for the existing system, I think we always assume that same framework (e.g. same RAN architecture) is applied to all RAN1/RAN3 features. Therefore, we don't see why for AI/ML, we need different RAN architectures for physical layer use cases and higher layer use cases. In addition, we don't think that the current RAN3 framework is only for inter-network-node without air-interface involvement, the function modules (e.g. data collection, model training, and model inference) defined in the RAN3 framework are exactly what we are talking here for physical layer use cases. If people really want to leave some room to see if there is any problem to do this, we can accept to just take it as baseline now, and we can further discuss if any modification really needed in the study phase. Therefore, we suggest to modify the last sentence in the justification as below:

For the study on AI/ML for air interface, the basic framework agreed for *FSNRENDCdatacollect*, as captured in section 4 of TR 37.817, should be taken **as baseline ~~into consideration for possible applicability.~~ Further modification is not precluded if identified as really necessary during the study phase.**

4.2 Candidate SID "Objectives"

Similar to the "Justification" section and based on the many inputs received without getting into one-by-one replies, the Moderator would like to propose the following revised text for further inputs during this final round of discussions.

Unfortunately, NWM does not support revision marks. Therefore, the Moderator is pasting here the clean version of the revised text and the revision marks with changes from the previous version can be seen at:

www.3gpp.org / ftp / tsg_ran / TSG_RAN / TSGR_94e / Inbox / Drafts / [RAN94e-R18Prep-08]

Table 7: Objectives section of candidate SID

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance, complexity, and potential specification impact.
Use cases to focus on:

Initial set of use cases for characterization and evaluation includes:

- CSI feedback enhancement, e.g., overhead reduction, improved accuracy
- Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction
- Positioning accuracy enhancements for different scenarios including, e.g., those with heavy NLOS conditions
- [RS overhead reduction]
- [Mobility]

Finalize representative set of use cases (from the initial set) for characterization and baseline performance evaluations
AI/ML model and description:

- Characterize the defining stages of AI/ML related algorithms:
  - Model generation, e.g., initiation/configuration, training (online/offline), model validation, model testing, as applicable
  - Inference operation, e.g., input/output, pre-/post-process, as applicable
- Identify various levels of collaboration between UE and gNB, e.g.,
  - No collaboration: implementation-based only AI/ML algorithms without information exchange
  - Various levels of UE/gNB collaboration targeting at separate or joint ML operation.
- Identify lifecycle management of AI/ML model: e.g., data collection, model deployment, model monitoring, model updating, and model transfer
- Data set for training/validation
- Identify common framework to characterize investigations related to AI/ML for Air-Interface
  - Consider the work done for *FS_NR_ENDC_data_collect* as and when appropriate

For the use cases under consideration:

1. Evaluate performance benefits of AI/ML based algorithms:

- Methodology based on statistical models (from TR 38.901 and TR 38.857 [positioning]), for link and system level simulations.
  - Extensions of 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
  - Field data to further assess the performance and robustness in real-world environments should be discussed as part of the study. User data privacy needs to be preserved.
  - Need for common dataset construction for training, validation and test for the selected use cases
  - Consider offline training vs. online training of models.
- KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations with

The Moderator would like to restate the fact that taking all the principles agreed in the ongoing RAN3-led SI on further enhancement for data collection *FS_NR_ENDC_data_collect* without a proper review of their applicability (many times RAN3-centric in their current form) is premature and unfounded. By the way, the TR referenced multiple times should be 37.817 and not 21.905. Similarly, reserving the term ”framework” to just the RAN3 view on their corresponding perception of AI/ML as part of the enhanced data collection study does not seem to be necessary, with all due respect to the work done on that project.

While the study is RAN1-led the early involvement of RAN2 and RAN4 WGs is seen as beneficial and necessary. However, some time should be given to RAN1 to improve its understanding of the problem to solve. In that sense, the Moderator suggests a late start of RAN2 and RAN4 involvement (e.g., 1-2 Quarters given to RAN1 to develop understanding before getting RAN2/RAN4 involved). Details TBD.

The Moderator thanks RAN2 chair comments and would like to recommend avoiding blind architecture options discussions. Early RAN2 involvement is very much needed for a complete understanding of enabling AI/ML over the Air-Interface. Input on meaningful RAN2 objectives is very much welcome.

Use cases have been heavily debated from the start of the discussions. The Moderator started with 3 use cases in this round of discussions because in Aug/Sept we could not agree with the 5 that were put on the Table (some companies wanted their favorite one added which were further increasing the list of use cases to consider). This is the only reason the Moderator decided to go back to the most popular use cases by far (as a response to some received questions).

In this current set of Objectives, the Moderator has put back in square brackets those two additional use cases (RS overhead reduction and Mobility) and has clarified the ”finalize representative set” bullet underneath so that we have final decision on whether to take all 5 or a subset for evaluations and characterizations. The original intent of that bullet was to discuss the additional use cases beyond the three which we had already identified.

Having said all that, the Moderator believes that the work on the various areas: use cases, AI/ML model/description, evaluations (methodology, assumptions, KPIs, etc) can be largely parallelized byfalling into possibly different RAN1 agenda points from the start. As a result, the Moderator does not see the need to have a two-phase study.

Please include your comments on the suggested revised ”Objectives” section for the candidate SID:

Feedback Form 12: Inputs on ”Objectives” section for candidate SID

1 – vivo Mobile Communication Co.

Thanks for the moderator’s efforts, we have the following comments. Pls also find a revised version of the SID in the draft folder based on our comments (https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_94e/Inbox/Drafts/%5
R18Prep-08%5D/DRAFT%20RP-212708%20SID%20on%20AI-ML%20for%20Air%20Interface%20RevMarks_v01_v

1. For use case to focus part, we prefer to have the following revision:

a) For CSI feedback enhancement part, we prefer to add CSI prediction back into the examples.

b) We prefer to remove the brackets over [RS overhead reduction] and [mobility], i.e., ~~[~~**RS overhead reduction**~~]~~**, ~~[~~**mobility**~~]~~**.** Based on level of support, these two cases should not be precluded before further

study. For the comments worrying scope of the discussion, it should be noticed that

i. For evaluation of RS overhead reduction would require much less effort to construct the data set compared to other use cases.

ii. For mobility, the major discussion may not directly be in the same discussion group as other use cases, thus the scope could also be well controlled.

c) The intention of the finalized representative set is to have a chance for more use cases. Thus it should be changed to: Finalize representative set of use cases (from the initial set **and with potentially new use cases**) for characterization and baseline performance evaluations. If only limited to be from the initial set, we don't understand the motivation to have a finalized set.

d) Is there any difference between **baseline performance evaluation** and **evaluation**? It seems that we use different terminologies for initial set and finalized set.

2. For AI/ML model and description part, we prefer to put the following in the main bullet: "Identify common framework to characterize investigations related to AI/ML for Air-Interface" as "**AI/ML model and description to identify common framework to characterize investigations related to AI/ML for Air-Interface"**.

3. For "assess potential specification impact" part, to align with what is described in the main bullet, the sub-bullet can be updated as following:

a) PHY layer ~~specification impact~~ **aspects** including…

b) Protocol ~~needs~~ **aspects** including…

2 – NTT DOCOMO INC.

Support the updated Objectives.

3 – Guangdong OPPO Mobile Telecom.

As for the use cases, we think RS reduction is well deserved in this study. we are fine to leave mobility use case out.

As for the data set, we think it is needed also for inference

As for the AI/ML generation, we think assumption on AI/ML model complexity is important from beginning. Otherwise we may spend more efforts to calibrate AI/ML models among companies.

please find other comments in the uploaded version2

4 – VODAFONE Group Plc

We agree with Vivo. We prefer to add CSI prediction back into the examples and not preclude RS overhead reduction or mobility.

5 – Keysight Technologies UK Ltd

We agree with the proposal. Although we would like to suggest removing **"late TBD start"** from RAN2 and RAN4 bullet points. We agree with the idea of delaying the start of the SI in RAN2 and RAN4, but we think timing info should be left out of this document and be considered as part of the TU allocation spreadsheet.

We would like to propose the following text changes:

- Identify lifecycle management of AI/ML model: e.g., data collection, **model training, model testing, model validation, model inference,** model deployment, model monitoring, model updating, and model transfer

- Data set for ~~training/validation~~ **training, testing and validation**

We would also like to add BS impact to RAN4 work by modifying the text as follows:

- UE **and BS** requirement**s** and testing framework**s** to validate AI/ML based performance enhancements and ensuring that UE **and BS** with AI/ML meet~~s~~ the existing minimum requirements

6 – ZTE Corporation

Use case

We agree with the moderator that there is no need to have a two-stage study. To have a balanced initial set of use cases is very important to ensure a careful and thoughtful study. Hence we support to have the original three use cases in the Rel-18 SI. The other two use cases under brackets can be left for future releases.

Further, we agree with the moderator that various areas can be put in different RAN1 agendas during the SI. Hence the evaluation and AI/ML model description should consider all the use cases in the initial set during the study, as most of the aspects are use case specific. Considering this, we suggest the following change to make this aspect clear.

Finalize the initial set of ~~representative set of~~ use cases ~~(from the initial set)~~ for characterization and baseline performance evaluations

Specification impact

For the RAN2 part, as we expressed in the inter-mediate round, what to study in RAN2 without clear RAN1 request is doubtful. It is clear the major functionality of the use cases and AI model LCM should be studied in RAN1. Usual RAN2 work like capability signaling design and RRC/MAC CE signaling to support UE/gNB collaboration can be done in future WI directly. It seems no need to study them first in RAN2 in an SI. Hence we suggest to revise the RAN2 bullets as follows.

- *Protocol needs including (RAN2 late TBD start)*
  - *Consider aspects related to ~~capability indication,~~ configuration procedures (training/inference), validation and testing procedures, and management of data and AI/ML model based on RAN1 input or request*
  - ~~*Collaboration level specific specification impact per use case including signaling design to support the collaboration identified in RAN1*~~

7 – AT&T

We can accept the proposed objectives assuming that brackets are removed for [RS overhead reduction] [Mobility] use cases.

8 – Deutsche Telekom AG

We are generally fine with the objectives.

For sub-bullet ”Identify lifecycle management of AI/ML model” ”data collection” should be deleted in the text as this is seen as part of the AI/ML framework, but does not belong to model LCM.

9 – Futurewei

In general, we are ok with the proposed "Objectives". A few comments:

- "AI/ML model and description":
  - "Characterize the defining stages of AI/ML related algorithms": it's not clear what "initiation/-configuration" means. If it's configuring model parameters, then it is related to model/algorithm design which should be left to vendor implementation. If it is to configure how to train the AI/ML model, then it belongs to model management. We suggest removing "initiation/configuration" from the description to avoid confusion. On the other hand, input/output and data pre-processing should be included in this stage as they may have specification impact, in addition to being included in the "Inference operation" stage.
  - "Data set for training/validation" should be modified to include testing, i.e., "Data set for training / validation / testing"
- "For the use cases under consideration":
  - "Evaluate performance benefits of AI/ML based algorithms"
    - "KPIs": suggest modifying the first bullet to "Performance and complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML **and/or** implementation-based AI/ML) baseline" as not necessary both baselines are feasible for all the use cases.
  - "Assess potential specification impact and aspects only for implementation, specifically for the agreed use cases and for a common framework"
    - The "only for implementation" wording used in this bullet is confusing; please remove it and just state "… for the agreed use case and for a common framework", which should be sufficient at this stage.
    - The sub-bullets under each RAN group, i.e., "PHY layer specification impact including (RAN1)", "Protocol needs including (RAN2 late TBD start)", and "Interoperability and testability aspects (RAN4 late TBD start)", are to be studied in detail in SI phase to determine the actual impacts. It is premature to specify them at this stage; thus, our suggestion is to either remove those sub-bullets or use a few of them as examples.
- The last sentence in the note "Typical AI model(s) can be provided for calibration" is not necessary and confusing as there are no definitions/descriptions for "typical AI model" and "calibration" so far. If such needs are raised during SI phase, they can be discussed at that time.

We would like to thank the moderator for coordinating and summarizing the discussions across companies.

10 – ROBERT BOSCH GmbH

We support the updated objectives

11 – CAICT

Thanks for moderator's great efforts and we are generally fine with the SID. There are two minor comments:

1 Use cases to focus on:

we believe channel estimation enhancements/RS overhead reduction and mobility enhancement could be discussed during the initial set use case and the bracket for these two use cases could be removed.~~[RS overhead reduction], [Mobility]~~

2 AI/ML model and description:

"No collaboration: implementation-based only AI/ML algorithms without information exchange" could be removed.

12 – InterDigital

Overall we support the proposed objectives. We would like to suggest the following updates:

- Note on the "Typical AI models can be provided for calibration" can be edited to "The need of Typical AI models for calibration can be studied' and moved under evaluation bullet.

13 – Qualcomm Incorporated

Overall, we support the proposal.

In particular, we prefer to retain the stated use cases as is, and having the final representative use cases as a subset from the initial set (i.e. not adding any additional use cases). Adding additional use cases would defeat the purpose of having selected initial use cases to contain the scope of the discussion.

We are OK removing the mobility use case, as it seems to overlap with beam management and RAN3 study.

14 – Apple Hungary Kft.

Overall we support the proposed objectives.

On use cases, mobility can be a RAN2 use case as well. RAN2 can determine whether to start the SI on mobility use case with/without RAN1 input. We also recommend remove the [ ] for RS overhead reduction.

15 – Intel Technology India Pvt Ltd

- we have concerns on scope if 5 categories of use-cases are considered for evaluation. Each category can have many sub use-cases for e.g. RS overhead includes DMRS/SRS/CSI-RS. we prefer previous version
  - Initial set of use cases ~~for characterization and evaluation~~ includes:
- we prefer to leave this for EVM discussions in RAN1, there is no definition of online/offline yet.
  - ~~Consider offline training vs. online training of models~~
- "Field data to further assess the performance and robustness in real-world environments should be discussed as part of the study" and "User data privacy needs to be preserved" can be separate bullets - they are not related.

16 – China Mobile Com. Corporation

Given the heavy workload, We hope 1-2 representative set of use cases can be selected for performance evaluations.

For whether and when to trigger model updating, it is related to the generalization capability of the AI/ML model. We think generalization capability should be considered as a KPI of the AI model. Hence we suggest to revise the bullet as follows

- Overhead, power consumption and memory storage associated with enabling respective AI/ML scheme, **as well as generalization capability** should be considered and documented.

17 – LG Electronics Inc.

Fine in general. Given the majority companies' interest, we also prefer to remove the brackets over [RS overhead reduction] and [mobility], i.e., **~~[~~RS overhead reduction~~]~~, ~~[~~mobility~~]~~.**

18 – CATT

Thanks for the effort. Please find CATT comments below.

- In our understanding, the initial set of use cases are only for companies to get better understanding on each use cases, but not for performance evaluation. The final representative set of use cases is a subset of the initial set. The following revision is suggested.

  Initial set of use cases ~~for characterization and evaluation~~ includes:

- We don't agree to put 'RS overhead reduction' and 'mobility' back in the list. RS overhead reduction or channel estimation is mostly implementation issue. It would not bring much value on the construction of AI/ML framework for 3GPP. Mobility use case shall be studied in the RAN3-led study or work item, if necessary.

- As we commented earlier, we prefer not to include online training in this study item due to its complexity.

  Model generation, e.g., initiation/configuration, training (~~online~~/offline), model validation, model testing, as applicable

  ~~Consider offline training vs. online training of models.~~

- It should be made clear that the analysis on the specification impact should only be performed on the agreed use cases in the final representative set. The following revision is suggested:

  Assess potential specification impact and aspects only for implementation, specifically for the agreed use cases in the final representative set and for a common framework:

19 – Lenovo (Beijing) Ltd

We have the following suggestions:

- **Use cases to focus on**. If all use cases in the "initial set" will be evaluated, the motivation to "finalize representative set" is questionable. Thus, we don't think it is feasible to include 5 use cases for full evaluation, as it can be quite huge effort to evaluate each of these use-cases and possible redundant work to assist the identification of a common AI/ML framework. Therefore, we prefer to keep the original three use cases, and any more case is for further study.
- **Field data**. As we commented in the Intermediate Round, because there will be extra and un-defined assessment work on the "field data" to be *good* or not, we suggest not to consider it within the evaluation methodology, or at least to discuss whether it is really needed and workable or not, such as " - **Whether** field data **are needed** to further assess the performance and robustness in real-world environments should be used **discussed** as part of the study."
- We think whether offline or online training can be indicated in the "AI/ML model and description" part for the use cases under considerations. Thus, we suggest delete it as "~~Consider offline training vs. online training of models~~" or just keep "Consider offline training at first step" if it has to be indicated.

20 – Xiaomi Communications

We can only accept the proposal if the bracket for RS overhead reduction and mobility is removed.

there are several sub-use cases in some use cases, and may be different evaluation methodology and data set are required for different sub-use cases. That is also a kind of increase of workload. We believe exploring application in more areas (e.g., improved channel estimation, mobility)is more benefical compared with exploring application in different sub-use cases.

Therefore, we think we should keep the RS overhead reduction and mobility. To control the work load, we can just focus on one typical application in each use case. For example, for the use case of CSI feedback enhancement, we could just focus on the overhead reduction

21 – Verizon UK Ltd

We are fine with proposed objectives as long as the brackets are removed for RS overhead reduction and mobility use cases.

22 – Rakuten Mobile

We prefer to remove square bracket on RS overhead reduction and mobility. For the rest of items, we are generally fine for the candidate SID Objectives.

23 – Ericsson LM

We are concerned that a large scope, involving many use cases, will stall discussions.

A "pre-study phase to select use cases" will necessarily involve parallel discussions and evaluations against state-of-the-art Rel 16/17 solutions (including preliminary specification impact analysis) to justify further study of those use cases. We, therefore, find it hard to motivate such a pre-study.

A mobility use case would have to be RAN2 led and require L1/L2 mobility (but that feature is planned for Rel 18). Hence, we believe adding mobility would further complicate the SI and risk stalling discussions.

Otherwise, we support the draft SID objectives.

24 – Fujitsu Limited

We have following suggestions:

1. Use cases to focus on: suggest to revoe the bracket of 'RS overhead', i.e., ~~[~~RS overhead reduction~~]~~

We think channel estimation (RS overhead reduction) can be taken as a fundamental use case of applying AI/ML to air interface. Since the basic functionality of AI/ML here is to use neural network to learn channel features although use cases are different. In addition, channel estimation is an easy use case to be simulated/evaluated. It is helpful to reach common understanding among companies on AI framework with lower efforts compared to other uses cases.

2. AI/ML model and description: suggest to add one bullet, i.e., - Conclude common terminologies for AL/ML related functions/procedures/interface which will be used in this SI.

Since AI/ML is a new area for RAN1/2/4, it would be better to have a common understanding on relevant terminologies for relevant AI/ML discussions in this SI.

3. KPIs: suggest to delete 'power consumption', i.e., Overhead, ~~power consumption~~ and memory storage associated with...

Since it is difficult to measure and quite hardware/software sensitive issue.

25 – Sony Group Corporation

Thank you for updating the objectives.

We are basically fine with the version. We think the study identifies potential use cases which means not only use cases from the initial set, buts also other potential use cases with a standards impact. The study should be about how (in terms of specification impacts) AI/ML can be used in the air interface, where the understanding of "how" is gained from the study of the selected use cases.

26 – Spreadtrum Communications

Generally we are fine with Moderator's proposal. No strong preference on whether to remove the bracket on 'RS overhead reduction' and 'Mobility' for the initial set of use case.

27 – Samsung Electronics Polska

Thank you for the updates. The general objectives looks good.

However, we have some concern on the **use cases part.** As we commented before, the listed use cases may require heavy AI training. For study item phase, we strongly request to at least study one use case exploiting lightweight AI, e.g., **real-time fast online training with limited reference symbols to compensate non-linear distortion for better supporting high order modulation scheme such as 256QAM/1024QAM.**

Besides, please find some other proposed updates on the objectives as below:

For beam management: beam selection accuracy is another side for evaluate beam management performance which is the same importance as overhead reduction. We propose the following changes:

o *Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction,* ***beam selection accuracy improvement***

For common KPIs: Common KPIs have nothing to do with collaboration levels at least for the user cases we ever mentioned. Common KPI may be relevant to where AI/ML scheme is implemented.

o *KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations* ~~***with different collaboration levels***~~ ***with different implementation ends****. Determine the use-case specific KPIs and benchmarks of the selected use-cases.*

Besides, we think **delay and hardware condition** should also be evaluated and documented for AI algorithm. Therefore, the following changes are proposed:

o *Overhead, power consumption,* ~~***and***~~ *memory storage,* ***computation delay and hardware conditions*** *associated with enabling respective AI/ML scheme should be considered and documented.*

For RAN 4 part□ better performance is also expected than only meets the existing requirements

§ *UE requirement and testing framework to validate AI/ML based performance enhancements and ensuring that UE with AI/ML meets* ***or exceeds*** *the existing minimum requirements*

Besides, the hardware device/chipset capacity for AI/ML operations is a fundamental for implementing the AI/ML schemes. RAN4 should consider if the requirement of capacity is necessary for securing AI/ML real implementation. Therefore, we propose to add the one sub-bullet for RAN 4 part:

§ ***Feasibility of UE requirements for the capacity of AI/ML operations***

28 – Nokia Corporation

We can accept the moderator's objective proposal. In our view the initial use cases should not have more details or examples before the actual technical studies.

We support Samsung's clarification proposal for the RAN4 part: *UE requirement and testing framework to validate AI/ML based performance enhancements and ensuring that UE with AI/ML meets **or exceeds** the existing minimum requirements.* The first part of the sentence already defines that AI/ML should provide performance gains, which should be visible in the RAN4 requirements.

29 – NEC Corporation

For the use cases, we suggest to include the two additional use cases. We are fine with the other parts.

30 – Continental Automotive GmbH

We agree with AT&T, we are supportive if the brackets for [Mobility] and [RS overheat reduction] are removed. Both are important goals.

31 – FGI

We are fine with the updated objectives.

32 – HUAWEI TECHNOLOGIES Co. Ltd.

1. Change "Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance" to "Study **~~the 3GPP framework for~~** AI/ML for air interface corresponding to each target use case regarding aspects such as performance". As we commented in the previous round, the definition of "3GPP framework" is not clear, thus to avoid the confusion better to delete it.

2. Make the following changes under "AI/ML model and description:", reasons are similar as what provided for justification above, we think the basic framework agreed for RAN3 SI should be taken as baseline.

Identify common **procedure ~~framework~~** to characterize investigations related to AI/ML for Air-Interface

~~Consider the work done for *FSNRENDCdatacollect* as and when appropriate~~

3. Make the following change for objective 2.

Assess potential specification impact and aspects only for implementation, specifically for the agreed use cases and for a common **procedure ~~framework~~**:

4. At least the bracket for RS overhead reduction should be removed. As we commented in the intermediate round, RS enhancement achieved broad interest also, just similar as the 3 use cases listed here, therefore no reason to only preclude RS enhancements. In addition, it is expected that the potential gain achieved for RS enhancements can be similar as other cases like CSI feedback enhancements also.

33 – MediaTek Inc.

We thank moderator's diligent efforts to consider companies' input and revise accordingly, we can see the improvement from last version, however, we do have following suggestions to make the plan better and to avoid unnecessary confusion and clarification later.

For use cases, we support the new use cases and suggest to assign responsibility explicitly

- CSI feedback enhancement, e.g., overhead reduction, improved accuracy **(RAN1)**
- Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction **(RAN1)**

- Positioning accuracy enhancements for different scenarios including, e.g., those with heavy NLOS conditions **(RAN1)**
- RS overhead reduction **(RAN1)**
- **RRM Mobility, e.g. prediction in time or frequency for robustness, interruption and overhead reduction (RAN2)**

With above clarification, we do not agree with Ericsson on the overlap with R-18 mobility enhancement, which aims to speed up mobility procedure and has nothing to do with AI/ML evaluation.

On common framework, although we also wish it could the case, it should be one of the outcomes after sufficient study on all use cases. Both ”common” and ”framework” are vague concepts for technical discussion, we would like to avoid such vague description to avoid misled discussion in WGs. Again, it only makes sense to discuss the possibility or the look of a ”common framework” after we confirm the gain and understand the architecture requirement for each use case.

Therefore, we suggest following changes:

1st suggestion

Suggest to modify ”Identify common framework to characterize investigations related to AI/ML for Air-Interface” –> ”**Identify common characteristics for framework investigation....**”

2nd suggestion, we welcome moderator’s comment on starting R2/4 work (the part not related to use case evaluation) at a later timing. This makes sense and we’d like to emphasize that general framework evaluation without substantial requirements, i.e. conclusion from use cases study, would definitely waste valuable WG time and should be avoided by all means, **and if it is the case, we cannot accept such planning.**

- PHY layer specification impact including (RAN1)
  - Consider aspects related to the specification of the AI Model lifecycle management, and dataset construction for training, validation and test for the selected use cases
  - Use case and collaboration level specific specification impact, such as new signalling, assistance information, measurement, and feedback
- Protocol needs including (**Except use case study, RAN2 only start following general evaluation after there is sufficient progress on use case study in RAN1**)
  - Consider aspects related to capability indication, configuration procedures (training/inference), validation and testing procedures, and management of data and AI/ML model
  - Collaboration level specific specification impact per use case including signalling design to support the collaboration identified in RAN1
- Interoperability and testability aspects (**RAN4 only start the work after there is sufficient progress on use case study in RAN1 and RAN2**)

  We also support Samsung’s revision suggestion.

5 Conclusions

5.1 Candidate SID ”Justification”

The Moderator would like to thank all the comments received in the final round of discussions. There is very broad support for the Justification section with some minor adjustments requested by a few companies.

Regarding the applicability of the basic framework agreed as part of the ongoing RAN3 project on enhanced

data collection (*FS_NR_ENDC_data_collect*), one company insists in taking it as the baseline while 28 companies are ok with the current version which states that the framework and principles agreed for *FS_NR_ENDC_data_collect* should be taken into consideration for possible applicability. Without the desire to ignore the comment, it is the Moderator's belief that the decision to take the basic framework in a related project led by a different WG as the the baseline requires first a proper assessment that has not been made. As said, many of those principles were made with RAN3 applications in mind as that is the scope of the project.

Based on the received comments, the Justification section is updated as follows:

Table 8: Justification section of the candidate SID

The application of AI/ML to wireless communications has been thus far limited to implementation-based approaches, both, at the network and the UE sides. A study on enhancement for data collection for NR and ENDC (*FS_NR_ENDC_data_collect*) has examined the *functional framework for RAN intelligence enabled by further enhancement of data collection through use cases, examples etc. and identify the potential standardization impacts on current NG-RAN nodes and interfaces*. In SA WG2 AI/ML related study, a network functionality NWDAF (Network Data Analytics Function) was introduced in Rel-15 and has been enhanced in Rel-16 and Rel-17. In this study, we explore the benefits of augmenting the air-interface with features enabling improved support of AI/ML based algorithms for enhanced performance and/or reduced complexity/overhead. Enhanced performance here depends on the use cases under consideration and could be, e.g., improved throughput, robustness, accuracy or reliability, or reduced overhead, etc. Through studying a few carefully selected use cases, assessing their performance in comparison with traditional methods and the associated potential specification impacts that enable their solutions, this SI will lay the foundation for future Air-Interface use cases leveraging AI/ML techniques. The goal is that sufficient use cases will be considered to enable the identification of a common AI/ML framework, including functional requirements of AI/ML architecture, which could be used in subsequent projects. The study should also identify areas where AI/ML could improve the performance of air interface functions. The study will serve identifying what is required for an adequate AI/ML model characterization and description establishing pertinent notation for discussions and subsequent evaluations. Various levels of collaboration between the gNB and UE are identified and considered. Evaluations to exercise the attainable gains of AI/ML based techniques for the use cases under consideration will be carried out with the corresponding identification of KPIs with the goal to have a better understanding of the attainable gains and associated complexity requirements. Finally, specification impact will be assessed in order to improve the overall understanding of what would be required to enable AI/ML techniques for the air-interface. For the study on AI/ML for air interface, the basic framework and principles agreed for *FS_NR_ENDC_data_collect*, as captured in section 4 of TR 37.817, should be taken into consideration for possible applicability.

As a result, the Moderator considers this modified Justification stable for inclusion into the draft SID for this project in RP-212708 for further discussion in RAN#94e.

5.2 Candidate SID "Objectives"

The Moderator would like to thank again the participants for their inputs. The following revised version should address all the comments received.

To the proposal to take the basic framework and principles from the RAN3-led study in

FS_NR_ENDC_data_collect, the same reply as stated in the Justification section applies here. Similarly, avoiding the use of the word ”framework” because it has been used somewhere else does not seem to create consistency issues or confusions. As a result, the Moderator has not made the proposed change to change it to ”procedure”.

Regarding, the additional use cases, the Moderator has removed the squared brackets for the two additional use cases added in this final round of discussions. Note, however, that the Moderator has some concerns with the extra work load that additional use cases will bring. For that reason, a mention has been made to possibly reduce the final representative set of use cases, from the initial set, while minimizing sub use cases.

== Start of Objectives section

Study the 3GPP framework for AI/ML for air interface corresponding to each target use case regarding aspects such as performance, complexity, and potential specification impact.

Use cases to focus on:

- Initial set of use cases includes:
  - CSI feedback enhancement, e.g., overhead reduction, improved accuracy, prediction [RAN1]
  - Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction, beam selection accuracy improvement [RAN1]
  - Positioning accuracy enhancements for different scenarios including, e.g., those with heavy NLOS conditions [RAN1]
  - RS overhead reduction [RAN1]
  - RRM Mobility, e.g., prediction in time or frequency for robustness, interruption and overhead reduction [RAN2]
- Finalize representative set of use cases (reduced from the initial set and minimizing sub use cases) for characterization and baseline performance evaluations

AI/ML model and description to identify common and specific characteristics for framework investigations:

- Characterize the defining stages of AI/ML related algorithms and associated complexity:
  - Model generation, e.g., model training (including input/output, pre-/post-process, as applicable), model validation, model testing, as applicable
  - Inference operation, e.g., input/output, pre-/post-process, as applicable
- Identify various levels of collaboration between UE and gNB, e.g.,
  - No collaboration: implementation-based only AI/ML algorithms without information exchange [for comparison purposes]
  - Various levels of UE/gNB collaboration targeting at separate or joint ML operation.
- Identify lifecycle management of AI/ML model: e.g., model deployment (initiation/configuration), model monitoring, model updating, and model transfer

- Data set for training, inference, validation, and testing
- Identify common notation and terminology for AI/ML related functions, procedures and interfaces
- Consider the work done for *FS_NR_ENDC_data_collect* as and when appropriate

For the use cases under consideration:

1. Evaluate performance benefits of AI/ML based algorithms:

- Methodology based on statistical models (from TR 38.901 and TR 38.857 [positioning]), for link and system level simulations.
    - Extensions of 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
    - Whether field data are needed to further assess the performance and robustness in real-world environments should be discussed as part of the study.
    - User data privacy needs to be preserved.
    - Need for common dataset construction for training, validation and test for the selected use cases
    - Consider adequate model training strategy and associated implications, e.g., offline training vs. online training of models.
- KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations. Determine the use-case specific KPIs and benchmarks of the selected use-cases.
    - Performance and computational complexity of AI/ML based algorithms should be compared to that of a state-of-the-art (non-AI/ML and/or implementation-based AI/ML) baseline
    - Overhead, power consumption (including computational), memory storage, and hardware requirements (including for given processing delays) associated with enabling respective AI/ML scheme, as well as generalization capability should be considered and documented.
- The need to define Typical AI model(s) for calibration shall be discussed as part of this study.

2. Assess potential specification impact, specifically for the agreed use cases in the final representative set and for a common framework:

- PHY layer aspects including (RAN1)
    - Consider aspects related to, e.g., the specification of the AI Model lifecycle management, and dataset construction for training, validation and test for the selected use cases
    - Use case and collaboration level specific specification impact, such as new signalling, assistance information, measurement, and feedback
- Protocol aspects including (Except use case study, RAN2 only start following general assessment after there is sufficient progress on use study in RAN1)
    - Consider aspects related to, e.g., capability indication, configuration procedures (training/inference), validation and testing procedures, and management of data and AI/ML model
    - Collaboration level specific specification impact per use case including signalling design to support the collaboration identified in RAN1

- Interoperability and testability aspects (RAN4 only start the work after there is sufficient progress on use case study in RAN1 and RAN2)
    - UE and gNB requirements and testing frameworks to validate AI/ML based performance enhancements and ensuring that UE and gNB with AI/ML meet or exceed the existing minimum requirements
    - Consider the need and implications for AI/ML processing capabilities definition

Note: specific AI/ML models are not expected to be specified and are left to implementation.

== End of Objectives section

The Moderator considers this modified Objectives above stable for inclusion into the draft SID for this project in RP-212708 for further discussion in RAN#94e.

A revision marked version w.r.t. the version discussed in the Intermediate round has been uploaded to the same folder:

www.3gpp.org / ftp / tsg_ran / TSG_RAN / TSGR_94e / Inbox / Drafts / [RAN94e-R18Prep-08]

6 References

[1] RWS-210659, “Summary of RAN Rel-18 Workshop”, RAN Chair.

[2] RP-211662, “Moderator’s summary for discussion [RAN93e-R18Prep-12] AI/ML”, Moderator (Qualcomm).

[3] RP-212657, “Additional Guidance on RAN Rel-18 Email Discussion during October 20th to 29th”, RAN Chair.