**3GPP TSG RAN WG2 Meeting #126 R2-24xxxx**

Fukuoka, Japan, May 20th– 24th, 2024

Agenda Item: 8.1.2.2

Source: Intel Corporation

Title: Report of [POST126][032][AI/ML PHY] LCM (Intel/Samsung)\_Phase 2

Document for: Discussion and Decision

# Introduction

This report provides a summary for the following post-meeting email discussion:

* [POST126][032][AI/ML PHY] LCM (Intel/Samsung)

Intended outcome:

Phase 1: Agreable definitions (Samsung)

Phase 2: Reach common understanding of reactive/proactive framework for applicable functionality.

Deadline: long

Deadline for providing initial comments for phase 2 is July 26th, 2024, 10:00 UTC.

Deadline for final comments based on rapporteur’s phase 2 summary is Aug 2nd, 2024, 10:00 UTC.

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

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| OPPO | Jiangsheng Fan | fanjiangsheng@oppo.com |
| Xiaomi | Xing Yang | Yangxing1@xiaomi.com |
| Futurewei | Chunhui (Allan) Zhu | chunhui.zhu@futurewei.com |
| NEC | Satoaki Hayashi | Satoaki-hayashi@nec.com |
| vivo | Boubacar Kimba | kimba@vivo.com |
| Apple | Peng Cheng | pcheng24@apple.com |
| Huawei, HiSilicon | Jun Chen | jun.chen@huawei.com |
| ZTE | Fei Dong | Dong.fei@zte.com.cn |
| Lenovo | Congchi Zhang | Zhangcc16@lenovo.com |
| Ericsson | Marco Belleschi | Marco.belleschi@ericsson.com |
|  |  |  |

# Phase 2 Discussion

During phase 1 discussion, following definitions of functionalities (with majority view) are summarized:

Supported functionalities refer to functionalities that UE can indicate by using UE capability signalling.

Applicable functionalities refer to functionalities that the UE is ready to apply for model inference.

Activated functionalities refer to functionalities already activated and performing inference.

In phase 2, we will mainly focus on the signaling framework for applicable functionality reporting. To facilitate the discussion, following assumptions are considered:

* The discussion in Section 2.1 and Section 2.2 first focuses on beam management use case. Whether the conclusion for beam management can also be applicable to positioning use cases will be discussed in Section 2.4.
* The applicability reporting is only focusing on model inference purpose first, according to the observation from companies’ feedback in phase 1 Q3.

Before discussing the signaling framework, rapporteur would like to first discuss the understanding of NW-side and UE-side additional condition.

As we discussed in NW-sided model, UE-side additional condition may refer to UE speed, scenario, hardware capabilities, etc. The understanding of UE-side additional condition for UE-sided model is the same as NW-sided model. However, it is not clear what is the exact meaning of NW-side additional condition.

RAN1 summarized NW-side additional conditions in summarized in R1-2405680 :

* Mapping relationship of Set A and Set B, including ordering to (a set of ID, or resource )
* Consistency of downlink spatial domain transmission filters corresponding to the beams in Set A and Set B.
* QCL assumption
* The order of model input and model output.
* between RS and Tx beams can be pre-defined.
* Transmission power
* UE distribution
* antenna height
* Deployment scenarios (e.g., ISD, Umi/Uma)
* ensure consistency across different cells.

Note that it’s not rapporteur’s intention to discuss what is considered as NW-side additional condition or the definition of NW-side additional condition in this discussion. However, considering RAN2 is focusing on signaling framework of proactive/reactive reporting, it’s good to understand how this information is reflected in RRC signaling, if NW-side additional condition needs to be signaled over the air interface.

Therefore, rapporteur would like to collect companies’ view on what is the NW-side additional condition first before discussing the signaling procedure.

#### Q0-1: What is the representative of NW-side additional condition of a functionality in RRC signaling (e.g. network supported functionalities, network supported AI/ML resource configurations of the supported functionality, etc)?

|  |  |
| --- | --- |
| **Company** | **Comment** |
| OPPO | In our view, NW side resource configurations can be divided into two categories:  Category A1(specified configurations used by UE for specific functionality but needs alignment with NW): the configurations configured to UE via dedicated signaling or broadcast signaling or default manner, i.e. Category A1 resource configurations can be known by UE via 3GPP specified method. For instance, BWP configuration, RB configuration, gap configuration and so on.  Category A2(specified configurations used by NW for specific functionality without UE involvement): the configurations are not configured to UE but specified via some spec e.g. NG, Xn, F1 and E1 spec. For instance, load info, interference coordination info and so on.  Category B1(NW implementation-based configurations which can impact inference performance for UE sided model): the configurations are not configured to UE, i.e. Category B1 resource configurations cannot be known by UE via 3GPP specified method but such configuration can significantly impact the inference performance of UE sided model. For instance, beam and Tx port mapping relationship, NW antenna shape, Antenna dip angle, height of gNB and so on.  Category B2(NW implementation-based configurations which has no impact on inference performance for UE sided model): the configurations are not configured to UE, i.e. Category B2 resource configurations cannot be known by UE via 3GPP specified method and such configuration has no significant impact on the inference performance of UE sided model. For instance, the algorism configuration for HO decision and so on.  [Mediatek] To clarify, if the additional condition on the network side involves only implementation configuration and does not impact the inference process, we could exclude this category from consideration as it remains entirely transparent to both the UE and the air interface. Is that correct?  In our view, when we’re talking about NW-side additional condition, it’s in relation with Category B1 resource configurations above. Because for Category A1 resource configurations, based on UE capability reporting, NW usually knows which specified data can be workable from UE point of view. As for Category A2/B2 resource configurations, it has no impact on inference performance for UE sided model, which is not related to UE side applicable functionality judgement.  More specific, from UE perspective, NW configured inference configuration includes two parts:  Part 1: Category A1 configurations, i.e. Normal specified functionality configurations, e.g. set A and/or set B resources for BM inference.  Part 2: Category B1 configurations, based on RAN1 progress, RAN1 has intention to introduce associated ID to implement Category B1 configurations, which can address the privacy concerns from NW vendor.  Based on above, we have the following definition for NW-side additional condition:   * **NW implementation-based configurations which can impact the consistency between training and inference for UE sided model**;   OPPO2：Thanks Rapp to give more info from RAN1 for NW-side additional conditions, but I would like to clarify the bullets listed here has no consensus by RAN1, RAN1 just lists all the possibility raised up by companies, some bullets even only supported by one company, so the bullets here are just for information to know all the possibility raised up by companies.  I was a little confused by the clarification from Rapp adding later, if we don’t touch what is considered as NW-side additional condition, how can we justify that UE capability signalling is not enough, anyway we should clarify the relationship between UE capability and additional condition reporting. We understand UE capability reporting+ additional condition reporting can tell NW the information needed to configure UE properly. For instance, UE reports its supporting two beam prediction functionalities: 4(input)-to-8(output) prediction and 8-to-16, based on UE capability signalling, at this stage (NW only obtains UE AI functionality capability), NW already knows which RS resources can be configured to UE side based on UE capability signalling, there is no need for NW to ask for the applicability of a specific RS resources again based on UAI, this is also the logic we follow for legacy non-AI feature, what NW does not know is the NW-side additional condition associated with UE sided model, because UE can know the NW-side additional condition associated with UE sided model only when the corresponding model is available at UE side, NW cannot get this dynamic info in advance without UE reporting. So NW should ask for such info (NW-side additional condition) via non-UE capability signalling, after this step, all info is ready from NW point of view, NW can do the configuration properly. |
| Xiaomi | We agree NW side additional condition is necessary to determine the functionality applicability. But we understand the exact meaning of NW side additional condition is up to RAN1. It’s not clear whether 3GPP would define the exact meaning of NW side additional condition. One example of NW side additional condition can be antenna pattern.  The general definition of NW side additional condition is already specified in the TR as following,  For an AI/ML-enabled feature/FG, *additional conditions* refer to any aspects that are assumed for the training of the model but are not a part of UE capability for the AI/ML-enabled feature/FG. It does not imply that *additional conditions* are necessarily specified. *Additional conditions* can be divided into two categories: NW-side additional conditions and UE-side additional conditions. Note: whether specification impact is needed is a separate discussion. |
| Futurewei | It is not very clear to us what “representative” means. To our understanding, additional conditions are the configurations under which a model was trained. Note although we are talking about functionality-based LCM, we all understand a functionality is always supported by one or more models. Therefore, NW-side additional condition are not “network supported functionalities”. These additional conditions are provided to the UE to ensure the NW and the UE are aligned on the conditions the model(s) is trained so that the UE/functionality can pick the right model for inference, for the current environment and configurations. |
| NEC | For BM-Case1 and BM-Case2 with a UE-side AI/ML model, the necessity and potential BM-specific conditions/additional conditions for functionality(ies) and/or model(s) are considered at least from the following aspects:  - information regarding model inference  - Set A / Set B configuration  - performance monitoring  - data collection  - assistance information  Among the above aspects listed in the TR, we understand consistency of Set B beams and Set A beams across training and inference would be a valid NW-sided additional condition. Followings are some examples:  ・Set size consistency for Set B, Set A: consistency in number of beams and/or associated resources for Set B and Set A, across training and inference  ・periodicity consistency for Set B, Set A: consistency in periodicity of beams and/or associated resources for Set B and Set A, across training and inference  ・relationship of Set A/ Set B (Set B is a subset of Set A or not): consistency in relationship of beams and/or associated resources for Set B and Set A, i.e, whether Set B is a subset of Set A, across training and inference |
| vivo | Tending to agree with Rapp to have general understanding on the NW-side additional conditions in RAN2 to facilitate the discussion on signalling procedure. The detailed additional conditions can be left to RAN1 to decide.  From our understanding, the NW-side additional conditions can be categorized as two types:  - **Type 1**: Beam characteristics, e.g., beam boresight direction (azimuth and elevation), 3dB beamwidth. This information can be agnostic to UE, thus can be implicitly indicated to UE as associated ID.  - **Type 2**: Set A/Set B related info, e.g., the beam index of set B. This information should be explicitly sent to UE for model inference. |
| Apple | 1) First, we suggest **RAN2 not to discuss definition of NW-side additional condition and detailed metrics of NW-side additional condition (e.g. set A and set B config):**   * As Xiaomi mentioned, its definition was already agreed in RAN1 and captured in TR 38.864. Any new definition from RAN2 perspective will confuse RAN1. * The list of metrics of NW-side additional condition from Rapporteur are only RAN1 feature lead summary of company contributions. As far we know, RAN1 didn’t make agreement for any of them. At this stage, we see no need for RAN2 to have duplicated efforts.   2) Please note RAN1 had made agreement to support associated ID at least in single cell in RAN1#117:  Working Assumption  Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent at least within a cell   * FFS: whether/how UE assumption can be applicable for multiple cells (including the feasibility study)   Based on it, we illustrate our understanding on how one NW-sided additional condition is indicated in RRC message in below figure with the following explanation:   * One “NW-sided additional condition” is associated to one inference configuration (e.g. resource of CSI set A) and/or one training configuration (e.g. resource of CSI set B) identified by the same associated ID.   + The UE trains separate AI/ML models with different dataset collected via training configuration identified by its associated ID (i.e. one associated ID->one training configuration->one AI model). * For one NW-sided additional condition” identified by one associated ID, the UE determines it is met if it identifies the same associated ID is indicated in applicability reporting configuration (i.e. consistency between inference and training is ensured).   + Please note that the UE can determine whether it is met with only associated ID (i.e. inference config is optional). Specifically, if only associated ID is provided, it is proactive reporting; If both associated ID and its corresponding inference configuration is provided, it is reactive reporting. * One “Supported functionalities” includes one or multiple “NW-sided additional condition(s)” because NW can have different training/inference configurations for one supported functionality   + For example, beam prediction in temporal domain is one supported functionality. NW may provide the following 3 NW-sided additional conditions for it:   Associated ID=1 identifies configuration of Set A / Set B under 2TX\*2RX MIMO;  Associated ID=2 identifies configuration of Set A / Set B under 4TX\*2RX MIMO;  Associated ID=3 identifies configuration of Set A / Set B under 8TX\*2RX MIMO; |
| Huawei, HiSilicon | The network side additional conditions are needed to be known to the network as part of the functionality applicability reporting. However, the details of what needs to be considered for the NW-side additional conditions are currently discussed in RAN1, so there is no need to repeat this discussion in RAN2. It may just be an associated ID as currently assumed for the BM case as in RAN1 agreement below:  At RAN1#117, it was agreed:  Working Assumption  Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent at least within a cell   * FFS: whether/how UE assumption can be applicable for multiple cells (including the feasibility study)   At RAN1#116bis, it was agreed:  Agreement  Further study, for the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, where the NW-side additional condition may at least impact UE assumption on beams of Set A/Set B:   * Opt1: Based on associated ID (Referring to AI 9.1.3.3)   + FFS on what can be assumed by UE with the same associated ID across training and inference   + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework |
| ZTE | The intention of NW additional conditions for the functionality is to ensure the scenario consistency between UE side model/functionality training and inference, which has been already discussed in RAN 1 for model identification and the following alternatives are provided:  1: Associated Id  2: Performance monitoring.  With above potential solutions, the consistency can be implied by either associated Id or reflected by the performance monitoring, and hence there is no need for RAN2 to discuss the detail NW additional conditions piece by piece from RRC signalling perspective, it is too complicated and time consuming for RAN2 to do that. Moreover, RAN2 is not the WG to do the simulation so that is not a good idea for RAN2 to discuss the NW additional condition in detail.  Instead of discussing the detail NW additional conditions, RAN2 need discuss how to determine one functionality as applicable based on the potential solution as shown below :   * Assuming the associated Id is used, UE and NW consider whether one functionality is applicable by comparing the UE additional conditions with the current UE status/settings and the NW additional conditions with the current NW settings.   Assuming the performance monitoring, UE and NW consider whether one functionality is applicable by evaluating the performance of the functionality. |
| Mediatek | As RAN2 discusses the signaling framework, we are specifically focusing on those NW-side additional conditions that are perceivable to the UE, either through explicit or implicit methods.   * Explicit Signaling: Here, NW-side conditions are clearly defined and communicated. This may encompass details like radio resource configurations, including sets A and B, and potentially the cell ID. * Implicit Signaling: In this approach, NW-side conditions are not directly described but are inferred. The concept of an associated ID, as discussed in RAN1, exemplifies this method. It serves as an indirect representation of NW-side conditions, facilitating consistency between training and inference phases without overtly signaling each condition.   Per the RAN1 agreement, the associated ID emerges as a tool for maintaining the consistency of NW-side conditions across the training and inference processes. It acts as a proxy for NW-side conditions, offering the advantages of safeguarding sensitive NW implementation details and minimizing signaling overhead on the air interface. In our understanding, at least the associated ID can be considered as one representative of the NW-side condition.  However, it remains to be clarified if associated IDs alone suffice to represent all NW-side conditions, thereby eliminating the need for explicit signaling of such conditions. This determination falls under the purview of RAN1. |
| Lenovo | Rapporteur has listed the options of NW side additional condition under RAN1 discussion, and RAN1 is considering using an “associated ID” to represent the NW side additional condition. When it comes to the possible RAN2 RRC signalling impact, it would be stage 3 details and we believe it will be easier to discuss after RAN1 conclusion on the definition of NW side additional condition. In addition, NW-side additional conditions (if exist) will be necessary to determine the applicability of the functionality to maintain consistency in training and inference. Whether and how it is signalled over the air interface would depend on who determines the applicability. |
| Ericsson | We agree with HW, Xiaomi that RAN2 does not need to discuss this topic at the moment. NW-side additional conditions are needed for determining the applicability of the AIML functionality. However, how to represent them, e.g. via associated IDs and the granularity of the IDs (e.g. per resource set), is under discussion in RAN1.  RAN2 only needs to focus on protocol-related aspects (i.e. proactive/reactive reporting, activation/deactivation of functionalities), we do not need to discuss at the moment the content of the RRC configuration for the AIML inference which can be left to RAN1.  We also would like to stress that the NW-side additional condition from RAN2 point of view can just represented by possible inference configurations that the gNB can provide to the UE, and that can be represented with the associated IDs. We do not need to spend further time on what are the details of such inference configurations. That is up to RAN1, and RAN2 can discuss at a later stage how to include all the necessary IEs into the RRC signalling. |

Furthermore, based on how the UE trains each functionality, different functionalities may have their corresponding different NW-side additional conditions.

#### Q0-2: Do you think 1) NW-side additional condition of the functionality supported by the UE needs to be signalled to the network? 2) If it is signalled to network, is it include as part of UE capability, or as part of other RRC signaling (other than UE capability)? 3) If it is not signalled to network, what is the expectation from companies?

|  |  |  |  |
| --- | --- | --- | --- |
| Company | **1) Yes/No** | **2) UE cap/other RRC signaling** | **3) Behavior if it’s not signalled to network** |
| OPPO | Yes with comments  This can be achieved via explicit or implicit way.  For explicit way, UE report info associated to NW-side additional condition if NW does not indicate filtering info associated to NW-side additional condition;  For implicit way, UE report Yes/No info associated to NW-side additional condition if NW indicates filtering info associated to NW-side additional condition. | Other RRC signaling (other than UE capability) is better.  Based on RAN2 previous agreement, UE should report functionality applicability only when the corresponding model(s) is available, but the model availability status can be changed in a short time, e.g. serving cell change, in this case, UE may need to update the reported functionality applicability based on the latest NW side additional condition. UE capability signaling is not suitable for such dynamic reporting procedure, so other RRC signaling (other than UE capability) is better for NW-side additional condition reporting. |  |
| Xiaomi | Yes,  We understand it’s essential for NW to know the NW-side additional condition of the functionality in following scenarios,  1) if NW is the node to determine functionality applicability,  2) if NW would like to modify the NW side additional condition to enable certain functionality applicable. | If the model/functionality is trained by UE, the associated NW-side additional condition is reported by UE.  If the model/functionality is trained by NW, the associated NW-side additional condition doesn’t need to be reported by UE.  If we focus on the first case, other signalling is preferred.  As defined in the TR, the additional condition is determined by the training data. If UE collects more training data, the associated NW-side additional condition may change. Therefore, legacy capability signalling is not applicable to carry such information. |  |
| Futurewei | Yes (in general) | Because some of the conditions may change dynamically, they should not be included in UE capability. In addition, they are called “additional” conditions for a reason; they are not part of UE capabilities. |  |
| NEC | Yes | We think it is reasonable to allow the UE to report its supported NW-side additional condition of the functionality.  We understand different cells may have different NW-side additional condition. However, from UE perspective, some “UE supported NW-side additional condition”  (for example, Set A and Set B configurations used for UE-side model training) could be static. We think such static “UE supported NW-side additional condition” can be included as a part of UE capability.  Moreover, including static “UE supported NW-side additional condition” as a part of capability can reduce AI/ML configuration latency and configuration overhead. |  |
| vivo | Yes, for Type 2 additional condition i.e., Set A/Set B related info. For example, NW may have to further assess the additional condition applicability. | Other RRC signaling |  |
| Apple | **No:**  It is sufficient for UE to **only reports applicable functionalities** **which meet all below 3 conditions**:  1) NW-side additional condition (i.e. inference config and training config have same associated ID).  2) UE-side additional condition (e.g. current left memory/battery resource is sufficient to do inference).  3) UE completes model training and model is available in device.  Because 2) and 3) can only be left to UE implementation, we don’t think NW can do anything with UE reporting only 1) as intermediate result. Thus, we think it is **not necessary to report NW-side additional condition.** | **No need to report NW-sided condition, only need to report applicable functionality**:  1) See our response in Q0-1. According to latest RAN1 agreement, the UE determines whether one NW-sided condition is met via NW indicated associated ID. Thus, NW already knows NW-side conditions, and it doesn’t make sense for UE to report the info which the NW already knows (and provided by NW).  2) Since model availability and UE-side condition are also needed to determine applicable, it is meaningless for UE to report this intermediate result (i.e. only NW-side condition) to NW because NW can’t derive UE-side additional condition and whether model training is complete. | According to latest RAN1 agreement, the UE behaviour: it determines whether NW-sided condition is met via checking whether same associated ID indicated by NW. |
| Huawei, HiSilicon | Yes.  The network needs such conditions to align the training and the inference, and then to decide whether it is applicable from the network side. | We think 2) depends on the training, and this may not be the same from one cell to antoher cell (for local associated ID), and hence it cannot be part of UE capability. |  |
| ZTE | It depends on RAN1 discussion on NW additional conditions. | As we mentioned in Q 0-1, in RAN1 discussion for model identification, the NW additional conditions consistency can be either realized by an associated Id or implicitly evaluated by performance monitoring.  Assuming that associated Id is adopt by RAN1:  The answer maybe YES , besides it is not precluded that the signalling including NW additional conditions (e.g. associated Id) from the NW to UE is also possible.  Assuming the performance monitoring is adopt:  It means the NW additional condition would never be exchanged between UE and NW, the consistency between training and inference can be potentially reflected by the performance monitoring of each functionality, then the answer is NO | Assuming the associated ID is adopt:  If there is no any associated Id related to one functionality is reported to the NW, NW then consider such functionality can not be configured for the inference. |
| Mediatek | **No for AI/ML functionality applicability report**:  Before we can address the question at hand, it is essential to clarify several key points:   1. Purpose of UE reporting NW-side conditions:  * One potential purpose could be to signal the availability of AI/ML functionality, potentially initiating a model transfer procedure from the network to the UE. However, this scenario appears to be outside the scope of the current question.  1. AI/ML Functionality Applicability:  * Another reason for reporting could be to assist in determining the applicability of AI/ML functionality. * This leads to the subsequent question: Who bears the responsibility for deciding the applicability of UE-side AI/ML functionality?   Upon review, we concur with Apple's perspective that the UE is tasked with deciding the applicability of AI/ML functionality. This decision is based on a combination of factors, including the UE's additional conditions, the availability of AI/ML models for the functionality, and the NW-side additional conditions.  This understanding implies that the network must first signal NW-side additional conditions, such as associated IDs, to the UE.  If the network-side conditions are indeed signaled using IDs, it stands no reason that the UE should provide these IDs to the NW in advance. Providing NW-side additional condition to UE would enable to make a more informed and accurate decision regarding the applicability of AI/ML functionality. | Agree with Apple.  The key issue to address is determining who is responsible for deciding the applicability of UE-side AI/ML functionality. Based on our analysis, it is more appropriate for the UE to make this decision. This is because the UE's additional (internal) conditions are more dynamic, and the availability of AI/ML functionality on the UE side must also be taken into account.  The method by which the UE communicates the applicability of AI/ML functionality to the network side needs further discussion. This could potentially be conveyed through a functionality ID, if necessary, or perhaps through a combination of the associated ID and other relevant information. | The network provides the NW-side additional condition to the UE. Then UE determines the AI/ML functionality applicability. |
| Lenovo | Yes with comment  We have agreed last meeting to take UE determining the applicability as the baseline. In this scenario, it could be beneficial for UE to report the supported NW-side additional condition for each applicable UE-sided functionality to NW, such that NW may consider the NW-side additional condition to make sure the functionality is applicable from NW side.  If companies agree supporting the scenario wherein NW determines the applicability for a UE-sided functionality, then probably UE will need to. | In phase 1, majority companies think UE may not have AIML model available yet when reporting the supported AIML functionality via UE Capability signalling. In this case, not sure how the UE could report any information related to the applicability of the AIML functionality, including the NW-side additional condition. Therefore, it seems more reasonable to report the supported NW-side additional condition later on after the AIML model is available. Also, if we consider more flexible scenario that UE could further download a new AIML model for the same functionality, the associated NW-side additional condition could change as well. Considering above, maybe UAI is more suitable to convey those “applicability related information” to NW. |  |
| Ericsson | We agree with Apple. RAN2 should just focus on the applicability reporting for an AIML functionality, and the applicability reporting may or may not contain the NW-side additional conditions, and potentially also other info (depending on RAN1 progress).  Hence, the UE might not always need to report the NW-side additional condition to the gNB. For example, the gNB as part of the inference configuration can provide one or more Set A/B and then the UE can just reply by indicating that the AIML functionality is applicable according to one of such indicated set A/B. No need to indicate further info.  Whether to provide or not the NW-side additional condition depends also on whether proactive or reactive reporting is adopted. So we suggest RAN2 focusing on the applicability reporting (proactive/reactive) and discuss instead related protocol aspects. | Other RRC signalling.  This information should be exchanged as part of the applicability reporting, for which we have already discussed at length (since the SI) that capability signalling cannot be used, because the NW-side additional conditions are possible radio configurations that can change dynamically depending on the network/gNB to which the UE is connected. |  |

## Proactive Reporting

It was agreed in RAN2 #125bis and #126 meeting that supported functionalities (may or may not have available models) are reported in UE capability, while the existing UE assistance information (UAI) procedure is used for applicable functionality reporting in proactive reporting.

|  |
| --- |
| **RAN2 #125bis meeting:**   1. Which AI/ML-enabled Features/FGs and functionalities are supported should be standardized. The details wait for RAN1’s progress. “supported” means that the UE is capable of supporting the functionality and doesn’t mean neccesarily that the UE has the model available. FFS what functionality refers to. 2. Supported AI/ML-enabled Features/FGs and supported functionalities are included in UE capability. 3. Support proactive reporting of UE-sided applicable functionality, e.g., the UE reports its applicable AI/ML functionalities via UAI message/LPP message.   **RAN2 #126 meeting:**   1. For BM use case, As a baseline the UE determines whether a functionality is applicable. Existing UAI framework is used at least for proactive reporting of applicable functionality. FFS reactive |

According to the above agreements and companies’ inputs to RAN2 #126 meeting, rapporteur summarizes below signaling framework as an example of proactive reporting for discussion:



**Step 1**: Network sends *UECapabilityEnqiry* message to initiate the procedure to a UE reporting its AI/ML supported functionalities.

**Step 2**: UE sends *UECapablityInformation* message to network, containing supported functionalities at the UE side.

**Step 3**: Network configures UE that it is allowed to provide its applicable functionalities.

**Step 4**: UE sends applicable functionalities to network upon change of applicable functionality/condition.

**Step 5**: Network sends inference configuration for the applicable functionalities to the UE.

**Step 6**: Start inference/monitoring based on network/UE activation/deactivation.

In rapporteur’s understanding, in existing UAI framework, UE can only send UAI if configured to do (basic behaviour with UAI). Hence, the network shall at least configure the UE to provide assistance information, i.e. configure UE whether it is allowed to do UAI reporting or not.

#### Q1-1. Do you agree network shall at least configure the UE via *OtherConfig* whether it is allowed to do UAI reporting before applicable functionality reporting?

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO | Yes | This is the basic logic as in legacy, we support this.  One more thing is that whether NW can configure something more, e.g. filtering info, to limit the reporting overhead in UL. Of course, this can be evaluated as the second step. |
| Xiaomi | Yes | Aligned with current UAI procedure. |
| Futurewei | Yes | Reasonable. |
| NEC | Yes | Follow current UAI framework. |
| vivo | Yes |  |
| Apple | Yes as baseline with comments | We are fine to use current UAI procedure as baseline of proactive reporting.  However, please note that existing UAI framework can only report UE’s **condition/information changes** **on serving cell (i.e. the UE is only required to monitor condition changes in serving cell in existing UAI framework)**. We are not sure whether existing UAI procedure can work for **proactive reporting of neighbour cells’ applicable functionalities reporting which requires the UE to monitor condition change in neighbour cells**. Please note that reporting applicable functionalities of neighbour cells need to be supported for Rel-19 AI/ML, at least including:   * Beam prediction for neighbour TRPs (within Rel-19 AI/ML based BM) * L3 measurement prediction for neighbour cells (within Rel-19 AI/ML based mobility).   To resolve the issue, we think RAN2 can:   * Either extend UAI framework to neighbour cell reporting (and it needs to put a new UE requirement to monitor condition change in neighbour cells). * or leave neighbour cell reporting to reactive reporting with different RRC message framework. |
| Huawei, HiSilicon | Yes | For Step 3, we think RAN2 can discuss whether the NW can also indicate some configurations to the UE, e.g. all/some functionalities within UE capability information, some associated IDs. However, this discussion is not urgent for now, and can be further discussed. |
| ZTE | It depends on RAN1 discussion regarding NW additional conditions | As we mentioned in Q 0-1, in RAN1 discussion for model identification, the NW additional conditions consistency can be either realized by an associated Id or implicitly evaluated by performance monitoring.  Assuming the performance monitoring is adopt:  Then proactive applicability reporting is not needed since UE need to do the performance monitoring to determine the applicability of each functionality.  Assuming the associated Id is adopt:  It can be a candidate procedure for proactive applicability reporting, in addition to the UAI, there can be another alternative, e.g. the RRCReconfiguation/RRCReconfiguationComplete also can be taken into account for the proactive applicability reporting. |
| Mediatek | Yes as baseline. | In step 3, the network can provide the NW-side additional condition in *RRCReconfiguration* message.  Another possibility is that UE can response the AI/ML functionality applicability directly in the *RRCReconfigurationComplete* message. |
| Lenovo | Yes | As in legacy |
| Ericsson | Yes | If UAI is used otherConfig should be configured as legacy. However, the content of otherConfig should be discussed, e.g. for which AIML functionality/ies the applicability reporting is requested, the radio configurations that the gNB can give to the UE. |

As agreed in RAN2 #126 meeting, applicable functionality is determined by UE as baseline.

|  |
| --- |
| For BM use case, As a baseline the UE determines whether a functionality is applicable. |

After the UE is configured to provide applicable functionality information, the UE first needs to decide which applicable functionalities should be included in UAI and report to the network. Furthermore, as discussed in [6] and [12], the network may also be able to determine whether a functionality is applicable or not, according to NW-side additional conditions (i.e. without sending NW-side additional conditions to UE).

According to companies’ inputs to RAN2 #126 meeting, rapporteur summarizes below options on how UE/network decide the applicable functionalities:

* **Option 1: Joint decision by UE and NW**

In this option, UE-side additional conditions are known by UE internally, and NW-side additional conditions are known at network-side internally. First of all, UE decides UE-considered applicable functionalities based on UE-side additional conditions (e.g. UE speed, scenario, hardware capabilities, model availability, etc). UE reports **UE-considered** applicable functionalities to network. Network further checks its NW-side additional conditions by implementation and decides the **final** applicable functionalities by providing the corresponding configuration in Step 5.

However, the UE may also need to include NW-side additional condition of the corresponding reported UE-considered applicable functionalities in Step 4. Example of Step 3-5 of Option 1 is shown as below:



* **Option 2: UE decision**

**UE decides** the applicable functionalities based on UE-side additional conditions (internally known by UE) and NW-side additional conditions. This option seems to be a special case for Option 2, where network sends NW-side additional conditions to UE in Step 3 in advance for applicable functionality determination. Example of Step 3-5 of Option 2 is shown as below:



* **Option 3: NW decision**

**NW decides** the applicable functionalities based on UE-side additional conditions and NW-side additional conditions. In this option, the UE needs to report its UE-side additional conditions to NW via UAI. In rapporteur’s understanding, for proactive reporting, network only needs to provide network configurations to functionalities with available models at the UE side. Therefore, in this option, rapporteur assumes UE only needs to report UE-side additional conditions of available models in Step 3, while the network doesn’t need to know the exact model used by the UE. However, the UE may also need to include NW-side additional condition of the corresponding reported UE-considered applicable functionalities in Step 4. Example of Step 3-5 of Option 3 is shown as below:



#### Q1-2. Regarding to applicable functionalities determination, which option do you agree? Please also list any missing information that you think needs to be transferred to UE or reported to NW in Step 3/4 for applicable functionality determination.

|  |  |  |
| --- | --- | --- |
| **Company** | **Option 1/2/3** | **Comment (Is there any other information needed? why it’s essential?)** |
| OPPO | Option1/2 | For Option1, functionality filtering info may also be optionally included in step3, i.e. other config, to limit the reporting scope from UE side. This filtering info is supported by many UAI features in legacy, better to keep the door open for now.  For Option2, we’re wondering whether the NW-side additional condition included in step3 is configured per functionality or per cell, this may impact the reporting format in UAI from UE side. Of course, this may somehow touch stage3.  Both Option1 and Option2 can be workable. For Option1, the scalability may be better than Option2 considering mobility scenario, for Option2, NW may configure the latest NW-side additional condition in other config again to trigger UE reporting updated UAI while Option1 may still survive after serving cell change as all supported NW-side additional reporting reported from UE will transfer to target cell during HO.  For Option2, the Yes/No info in UAI may save some overhead compared to Option1.  Option3 is not workable from our side unless UE reports both UE side and NW-side additional condition and then NW make the decision. If UE only reports UE side additional condition, we don’t know how NW can know NW-side additional condition associated with UE side functionality. |
| Xiaomi | Comments | Option 3 is not preferred since it’s difficult to standardize and report all UE side additional condition. Many factors are up to implementation.  The key difference of option 1 and 2 is whether UE can obtain the NW side additional condition. It’s still under RAN1 discussion whether NW should provide NW side additional condition to UE.  We understand UE shall determine functionality applicability at least based on UE side additional condition. If UE can obtain the NW side additional condition, UE shall also consider the NW side additional condition. It’s unnecessary for UE to ignore the available NW side additional condition.  In short, UE shall determine the functionality applicability based on UE side additional condition and NW side additional condition if available. Option 1 and 2 can be merged. |
| Futurewei | Comments | We think it is important to separate the discussions for NW-side functionality and UE-side functionality; each may prefer different options. For example, for UE-side functionality/model, Option 2 may be preferred while for NW-side functionality/model, Option 3 may be preferred. |
| NEC | Option 2 | Phase 1 has the agreement that “Applicable functionalities refer to functionalities that the UE is ready to apply for model inference.”, therefore we think Option 2 is a straightforward solution.  Option 1: Regarding the NW-side additional conditions, if there are no information from NW side, seems the UE shall report UE-considered applicable functionalities and all corresponding UE-considered NW-side additional conditions which may not match with the current gNB additional conditions at all.  Option 3: This option against the baseline we agreed at RAN2#126,  As a baseline the UE determines whether a functionality is applicable.  Moreover, it is not clear that whether NW has full knowledge on the UE-side additional conditions to make a proper determination of applicable functionalities. |
| vivo | Option 1/2 | For step 3, the NW-side additional conditions can be broadcast as system information, e.g., associated ID |
| Apple | Option 2 | According to latest RAN1#117 agreement, NW indicates NW-side additional condition(s) to the NW via associated ID as we illustrated in Q0-1, and the UE determines whether it is met via whether identifying inference config and training config with same associated ID. This is aligned with option 2, and we don’t think option 1 and option 3 can work:   * Option 1:   + We fail to understand why “NW-side additional conditions” are explicitly reported in step 4. Please note that NW-sided additional condition is notified by NW to the UE in forms of associated ID. So, NW already knows such info. Then, why does the UE need to report the information which the NW already knows (and provided by NW)?   + Because UE-sided additional condition is not reported to NW in option 1, we don’t think NW can determine whether one functionality is applicable or not. Thus, option 1 is incomplete solution. * Option 3:   + **This solution can work only if the UE is specified to report UE-sided additional condition to NW.** However, as they may include privacy information and UE proprietary implementation, we strongly object to specify reporting of UE-sided additional conditions to NW.   + As NEC mentioned, this option is against the baseline we agreed at RAN2#126 (i.e. the UE determines whether a functionality is applicable). Considering Rel-19 AI/ML discussion is controversial, we should focus on baseline at this stage.   + Please note that reporting UE-side additional information is not in scope of Rel-19 WID, and RAN1 don’t study the detailed metrics of UE-side additional information. |
| Huawei, HiSilicon | Option 1, 2 | For option 1, UE just checks applicability by UE-side additional conditions, and then report the applicable functionalities to NW. NW will check this input with its associated ID information, and do the functionality management decision. This option is reasonable.  For option 2, NW can provide some functionalities/associated IDs to UE, so that the UE will use this information to check. Compared with option 1, the UE has more information (i.e. associated ID), and thus it can do the filtering. This option is also reasonable, as it could reduce some signalling in Uu. For example, if the UE has functionality for associated ID #1, #2, and #3, but the current NW only supports ID#2, then option 2 could let UE report #2.  For option 3, we do not see the usefulness. For UE-side additional conditions, the NW does not need to know the details, and instead they are just used by UE side. In other words, the necessity of UE reporting UE-side additional conditions to NW side for UE sided model should be clarified first. |
| ZTE | Comments | For option 1, it is too early to decide the contents of each procedure step since, in the Question 0-2, rapporteur asks what kind of RRC signalling could be used for reporting NW additional conditions from UE to NW (e.g. UE capability or RRC signalling other than the UE capability). Assuming the UE capability would be used for carrying the NW additional conditions (e.g. associated ID), then in option 1, there is no need to carry the NW additional conditions in the UAI, it is because, in this case, NW can filter the functionality reported in the UE capability by considering the respective NW additional conditions, and then make UE report the applicability of the **interested functionalities**(e.g. the NW additional conditions of those functionalities is considered as apt to the current NW settings) , in this sense, there is no need to have the NW additional conditions again in the UAI message.  For option 2, it illustrates that the NW additional conditions shall be reported from NW to the UE which has not been discussed in the above questions (e.g. above questions only mention the case that the NW additional conditions reported from UE to NW, and the NW additional conditions reported from NW to UE is NOT mentioned).  Regarding the option 3, it is as similar as the option 1 if the NW can do the filter according to the UE capability and the NW additional conditions are reported in UE capability, then NW can make the decision according to the received UE additional conditions, but whether or what kind of the UE additional conditions are needed shall be discussed in RAN1 not RAN2, in this sense, the option 3 shall be pending until RAN1 has confirmed the UE additional conditions is needed for NW to identify the UE side functionality. |
| Mediatek | Option 2 | Just as commented in Q0-2, UE is responsible for deciding the applicability of AI/ML functionality. This decision is based on a combination of factors, including the UE's additional conditions/internal conditions, the availability of AI/ML models for the functionality, and the NW-side additional conditions.  Under Option 3, the UE is expected to report both UE-side additional conditions/ internal conditions and the availability of AI/ML functionality. However, the UE-side additional conditions/internal conditions relevant to model training and inference are largely tied to the UE's proprietary implementation. Specifying these conditions would be impractical due to concerns over proprietary information, competitive differentiation, product strategy, and the potential complexity it would add to standardization efforts.  In the case of Option 1, it remains unclear how the UE would ascertain the applicability of AI/ML functionality without any preliminary information from the network. If the decision is based solely on the UE's internal conditions and the availability of AI/ML models, there is a high likelihood that the UE would deem all available AI/ML functionalities as applicable. Therefore, the usefulness of such information for the network to make a final determination on the applicability of AI/ML functionality is questionable.  Option 2 facilitates a more informed and precise decision-making process regarding the applicability of AI/ML functionality by the UE, assuming network provides the NW-side additional conditions to the UE. |
| Lenovo | Option 2 as baseline agreed last time  Option 1/3 are also possible | In the last RAN2 meeting, RAN2 agreed to support UE determining the functionality applicability as the baseline, which is Option 2.  In the meanwhile, we believe Option 1 or 3 are also possible. Eventually it will depend on the required input (e.g., NW/UE side additional condition) and if the value of them can be transferred over air interface. |
| Ericsson | Option 1 as baseline  Option 2 to be discussed together with the reactive approach | All the options are in fact the same thing, and we do not need to distinguish about UE decision vs joint UE-NW decision. In all the options, it is always the network that decides the configuration, and it is always the UE that checks its own applicability conditions. The issue is more the content of the messages in the various steps.  Additionally, we note that the proactive and reactive approach should coexist, and they should not be considered necessarily as alternative approaches, i.e. the NW can provide both a reactive configuration and a proactive configuration. In such a case, the proactive approach based on UAI can be used to inform the NW about changes in the applicability conditions considering the reactive configuration. This aspect should be considered when evaluating the signalling details of the proactive options, since the inference configurations (including the NW-side additional conditions) do not need necessarily to be conveyed via the OtherConfig.  Option 1 is more aligned with the current UAI framework. In the current UAI framework, the otherConfig carries minimum set of information just to enable the UE to report its preferred/recommended configurations to the gNB. Hence, option 1 is more aligned with this framework, i.e. the gNB enables the UE in the otherConfig to send UAI for one or more AIML functionalities, and in the step 4, the UE responds with the applicable functionalities and related NW-side additional conditions. Then in step-5 the gNB provides the needed AIML inference radio configuration. Option 2 can also be considered, but in that case we are not sure that the inference configurations including the NW-side additional conditions should be signalled as part of the otherConfig. The NW can just configure the UE according to the reactive approach and then the UE can use the UAI to inform the NW about the changes in the applicability considering the configurations included in the reactive configuration. |

#### Q1-3. Any other comment on the above signaling flow of proactive reporting?

|  |  |
| --- | --- |
| **Company** | **Comment** |
| OPPO | See answer in Q1-2. |
| Xiaomi | Option 1 and 2 can be merged. UE shall first determine the functionality applicability based on UE side additional condition and NW side additional condition if available. After report, NW can further choose the functionality based on NW side additional condition, which is not provided to UE. |
| NEC | For activating UE-sided model, at least two RRCReconfig are needed, the first one for applicable functionality UAI report and the second one for full AI/ML configuration. Not sure are there any concerns on latency. |
| vivo | Option 1 and 2 can be merged, for example in case NW side additional conditions are available and NW can include them in step 3 of Option 1, i.e., along the configuration to allow UE performing UAI reporting. |
| Huawei, HiSilicon | Agree with Xiaomi that Option 1 and 2 can be captured in one procedure. For example, if the NW indicate associated ID to UE, the UE checks both UE-side additional conditions and the indicated associated ID for generating applicable functionalities; otherwise, the UE only checks UE-side additional conditions for generating applicable functionalities and reports them together with the supported associated IDs (here it is assumed that the UE has not reported associated ID via UE capability signalling). |
| Ericsson | As commented above, we agree that option 1 and 2 are in principle the same thing, but the network actions might be different. Additionally, the reactive approach should be also considered when evaluating UAI-related signalling. For example, in option 2, the step 5 might not be needed, i.e. if the UE was already provided in the reactive configuration with the possible AIML inference configurations (including the NW side additional conditions), then in step 4 the UE can just signal the AIML inference configuration that is applicable, and then in step-5 the gNB can just activate the model without any further configuration. So at this stage we do not need to merge them, we just need to better clarify the signalling needed.  Further comments:  Option 1/2 – step 3: It should be clarified that the otherConfig may include the AIML functionality for which the applicability conditions should be evaluated by the UE (and for which UAI should be reported)  Option 1/2 – first box: it should be clarified that the checking of the applicability functionalities should be done on the basis of the OtherConfig configured by the gNB  Option 1 – step 3. It should be further discussed whether OtherConfig should be used. As commented above, UAI is typically used by the UE just to provide its recommendation/preference without specific inputs from the gNB. Here instead, it seems that the otherConfig should provide extra radio configuration information, that could be provided also via the reactive approach. We suggest further evaluating the option 2 together with the reactive approach progress |
|  |  |
|  |  |

## Reactive Reporting

According to companies’ inputs to RAN2 #126 meeting, it seems companies’ understanding on reactive reporting is whether the applicable functionality reporting is considered as a response to network configuration. The network can either provide AI/ML resource configuration for multiple functionalities or some initial query (with some initial configuration) of applicable functions to UE. UE then reports applicable functionalities in response, either indicating acceptance of certain configurations or request for detailed network configurations for applicable functionalities.

From signaling procedure point of view, rapporteur summarizes below example for reactive reporting. The overall signaling procedure is similar to proactive reporting. However, there are some differences in terms of signaling content and detailed information, for example, the network needs to provide configurations to UE before UE reports applicable functionalities.



**Step 1**: Network sends UECapabilityEnquiry message to initiate the procedure to a UE reporting its AI/ML supported functionalities.

**Step 2**: UE sends *UECapablityInformation* message to network, containing supported functionalities at the UE side.

**Step 3**: Network provides network configurations and initiates UE to report its applicable functionalities.

**Step 4**: UE sends applicable functionalities to network.

**Step 5**: Network sends updated inference configuration for applicable functionalities reported in Step 4 to the UE. (see Q2-6)

**Step 6**: Start inference/monitoring based on network/UE activation/deactivation.

Following agreement was reached in RAN2 #125bis meeting:

1. Support reactive reporting of UE-sided applicable functionality. The NW configures AI/ML functionalities via RRC/LPP message. FFS what the configuration contains. FFS how to report applicable functionality and what is applicable functionality

In rapporteur’s understanding, network can provide configurations (e.g. beam resource configuration of Set A and Set B) for supported functionalities in Step 3.

#### Q2-1. In Step 3 of reactive reporting, do you agree that network can provide some configurations (e.g. AI/ML beam resource configuration of Set A and Set B) for functionalities? Please also provide an example of configuration for functionalities other than AI/ML beam resource configuration, if applicable.

NOTE: NW-side additional condition will be discussed in Q2-2.

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO | No | As analysed in Q0-1, NW will definitely know which specified configuration, i.e. Category A1 resource configurations in Q0-1, is or is not applicable to UE side based on UE capability reporting, this is the logic we always follow in legacy, what is the strong motivation to break this principle for AIML? why NW should ask again when already known via UE capability reporting? What NW does not know is the NW side additional condition associated to UE side functionality, that’s why we need NW side additional condition reporting from UE to NW.  Based on above, we don’t think this kind of reacting method is workable.  From our side, reacting method have another understanding, like proactive method listed above, NW configure other config via RRCReconfiguration message, UE responses with RRCReconfigurationcomplete message including applicable functionality, the three options introduced for proactive method above is also valid for this reacting method. |
| Xiaomi | Yes with comments | This is related to the granularity of functionality, which is under RAN1’s discussion. It’s up to RAN1 whether the beam resource configuration of set A and B can identify a functionality. If functionality is agnostic to configuration of set A and B, there is no need to provide such configuration to trigger reactive report.  NW can already obtain the supported functionality via UE capability signalling, as concluded in phase 1. NW may just indicate the ID of the supported functionality. And UE response with the applicability of the indicated functionality. |
| Futurewei | Yes. | NW-side additional conditions/configurations will help the UE to determine its available functionality. |
| NEC | Maybe Yes | We understand the rapporteur intention here is to provide some configurations related to NW-side addition conditions. If Set A / Set B configurations can be considered/defined as NW-side additional conditions, then “AI/ML beam resource configuration of Set A and Set B” can be used to present NW-side additional condition related information. Based on this understanding, we think the answer may be “Yes”.  Basically, we think Step 3 can only include configurations related to NW-additional condition (e.g., configurations which used to present NW-additional conditions). Full configuration of the AI/ML, e.g., configuration for inference and/or performance monitoring, etc. of the applicable functionality should be provided after receiving UE applicable functionality reporting (i.e., Step 5). |
| vivo | Yes, with Type 1 i.e., Beam characteristics, no for type 2 i.e., Set A/Set B related info | - Type 1: Beam characteristics, e.g., beam boresight direction (azimuth and elevation), 3dB beamwidth. This information can be agnostic to UE, thus can be implicitly indicated to UE as associated ID.  - Type 2: Set A/Set B related info, e.g., the beam index of set B. This information should be explicitly sent to UE for model inference. |
| Apple | Yes with comments | In our understanding, whether inference configuration is provided together with associated ID in same *RRCReconfiguraiton* is one of the key differences between proactive reporting and reactive reporting:   * Proactive reporting:   + **Only associated ID(s)** are included *OtherConfig* of *RRCReconfiguration* (i.e. its corresponding inference configuration like set A configuration is not provided)*.*   + Since inference configuration is not provided, the UE can perform inference only after NW provides inference configuration with another message (i.e. inference config is provided **after** applicable functionality reporting). * Reactive reporting:   + **Both associated ID(s) and their corresponding inference configuration** are included *RRCReconfiguration.*   + Since inference configuration is already provided, the UE can **perform inference immediately for the functionalities which are determined as “applicable”**, without need to wait another message from NW (i.e. inference config is provided **before** applicable functionality reporting). |
| Huawei, HiSilicon | See comments | We are open to discuss "some configurations".  Firstly, we think there should be at least some examples for the configurations. Secondly, we suggest to also discuss the UE behaviours upon getting such configurations from NW side. As mentioned by some companies, if NW can include configurations related to NW-side additional conditions (e.g. associated IDs), the intention should be let UE do the filtering. Then we think it is one example.  And then, we could discuss the necessity of such configurations. |
| ZTE | May be No, it depends, see the comments | As we mentioned in Q 0-1, in RAN1 discussion for model identification, the NW additional conditions consistency can be either realized by an associated Id or implicitly evaluated by performance monitoring.  Assuming the associated Id is adopt:  In this assumption, We think it is not a valid case for which the UE send the applicability reporting reactive to the RRCReconfiguration including the functionality configuration(e.g. Set A and/or Set B), in our understanding, NW shall be aware of the applicability of each functionality via associated Id before configuring it to UE (e.g. via proactive applicability reporting), otherwise, the RRC signalling to configure the useless AI/ML functionality would be waste due to such blind manner which seems not necessary.  Assuming the performance monitoring is adopt:  In this assumption, we think it is a valid case for which the UE send the applicability reporting reactive to the RRCReconfiguration including the functionality configuration(e.g. Set A and Set B beam resource configuration), and UE start executing the performance monitoring for those preconfigured functionalities and report the applicability of each functionality to NW after a time period of performance monitoring. |
| Mediatek | Yes with comment | The difference between proactive and reactive reporting lies in the timing of coordinating AI/ML functionality applicability relative to the inference configuration. In reactive reporting, the network provides the inference configuration to the UE before the coordination of AI/ML functionality applicability.  However, given that proactive reporting allows the UE to precisely indicate AI/ML functionality applicability before the inference configuration is set, the necessity for reactive reporting needs to be clearly justified.  One potential reason for having reactive reporting is its ability to accommodate dynamic changes in the UE's internal conditions, which may influence the applicability of AI/ML functionalities after the initial setup.  Another potential reason for having reactive reporting is to report the availability of the AI/ML functionality, which may trigger model transfer from the network to the UE if model transfer is supported. |
| Lenovo | No? see comments | First, it would be helpful to clarify the scenario, is it about UE determining the applicability (Option 2 in Q1-2).  If the intention of this question is asking if NW should provide some configuration for functionality before UE determines the applicability, then the answer would be yes. And example of the configurations would be those useful and related to applicability determination, e.g., related to NW side additional condition such as the Set A and Set B association discussed by RAN1.  But if the intention of this question is asking if the related configuration must be provided in the same request message that triggers the UE applicability report, the answer would be no. In our view, it can be provided in any regular *RRCReconfiguration* message beforehand, and not necessarily coupled with the trigger of UE applicability report. |
| Ericsson | Yes | The Step 3 in the reactive reporting can include one or more candidate radio inference configurations including the associated IDs that the UE can use for the inference and also the AIML functionalities that are requested by the gNB. Then in step 4, the UE can respond indicating which (if any) of these indicated radio configurations make the AIML functionality applicable, or it can also indicate other radio configurations (including the NW side additional conditions), e.g. if none of the inference configurations indicated by the gNB in step 3 are applicable. Related to step-5, we believe that it is optional. For example, if in step-4, the UE indicates that a certain inference configuration is applicable, then step-5 can just be an activation command, or simply the UE can activate/apply straight away the inference configuration if that is applicable for the AIML functionality (no need in this case for any activation)  Related to the question from the rapporteur to “provide an example of configuration for functionalities other than AI/ML beam resource configuration”, we believe that this is left to RAN1, and RAN2 should just discuss protocol related aspects, without digging into the content of the “inference configuration”. |

According to UE capability reported by the UE, network can provide network configuration (discussed in Q2-1) to all/subset of supported functionalities.

Based on current NW-side additional conditions, network can filter a list of functionalities from supported functionalities according to network implementation, and then provide configurations for those NW-considered applicable functionalities. Alternatively, similar as proactive reporting, in addition to network configuration (discussed in Q2-1), network can provide NW-side additional condition explicitly in Step 3, and let UE decide applicable functionality based on both NW-side and UE-side additional condition.

In the end, considering this configuration is provided by network based on UE supported functionalities, in rapporteur’s understanding, all of the functionalities configured by network in Step 3 may or may not have an available model, and may or may not be applicable at the UE side.

#### Q2-2. In Step 3 of reactive reporting,

**1) is there any other information (e.g. NW-sided additional condition) you think the network can provide except AI/ML resource configuration of NW-considered applicable functionalities (i.e. subset of supported functionalities based on NW-side additional condition)**

**2) those functionalities may or may not have an available model, and may or may not be applicable at the UE side?**

|  |  |  |  |
| --- | --- | --- | --- |
| Company | **1) Yes/No** | **2) Yes/No** | **Comment** |
| OPPO |  |  | See answer in Q2-1 |
| Xiaomi | Yes | Yes | We agree with rapp NW may provide the NW side additional condition to UE. UE can response with the applicable functionality under the corresponding NW side additional condition. Such information is useful for NW to decide the appropriate NW configuration and activated functionality. |
| NEC | Maybe No | Yes | We are a bit confused with the relationship between “AI/ML beam resource configuration of Set A and Set B” in Q2-1 and “AI/ML resource configuration of NW-considered applicable functionalities” here.  If the former one is just an example of the latter wording, we have no idea on any other information besides “AI/ML resource configuration of NW-considered applicable functionalities”.  Step 3 only provides NW-sided additional condition, therefore, those functionalities may or may not have an available model, and may or may not be applicable at the UE side. |
| vivo | Left to RAN1 | Yes |  |
| Apple | Yes (associated ID) | Yes | As we responded in Q2-1, NW-sided additional conditions are always provided to the UE in the form of associated IDs, irrespective of proactive reporting or reactive reporting. Thus, we think associated ID are needed to be provided. |
| Huawei, HiSilicon | Yes | Yes | For 1), as we commented for proactive reporting, the UE may only report the functionalities for NW-side additional conditions requested by NW. |
| ZTE |  |  | See our answer in Q2-1. |
| Mediatek | Yes. (i.e., NW-side additional condition, details left to RAN1) | Yes with more clarification |  |
| Lenovo | Need clarification | No, there may be problematic if the model is not available at that moment | 1) need to clarify the scenario of applicability determination it sounds like a variation of Option 1 in Q1-2. Maybe we should discuss first is Option 1 or any variation of Option 1 can be agreeable, then the same solution can work for both proactive and reactive.  2) as we raised in phase 1 discussion, if we consider the case that the model is not available at the moment NW triggers the applicability report from UE, it is difficult for NW to provide proper configurations/information that is related to the applicability of a functionality. |
| Ericsson | Yes | Yes | Question 1): The gNB when providing the inference configurations it should include the necessary associated IDs (NW-side additional conditions) to aid the UE to determine the applicability. From RAN2 pov, at least the following info should be included in step 3:   * Inference configurations (including the associated IDs/NW side additional conditions) * The AIML functionalities of interest for the NW.   Question 2): The AIML functionalities requested by the gNB in step 3 may or may not be available at the UE, because the NW cannot know beforehand (e.g. from capabilities) whether the UE really has a trained model for the request functionality |

Then UE can decide the final applicable functionalities based on its UE-side additional condition:



#### Q2-3. Do you agree UE reports final applicable functionalities (applicable based on both UE and NW side additional condtion) in Step 4, as a response to Step 3?

NOTE: Non-applicable functionality with available model reporting in Step 4 will be discussed in Q2-7.

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO |  | See answer in Q2-1 |
| Xiaomi | Yes with comments | We suggest to remove the ‘final’. Because NW may further determine the applicable functionality based on NW side additional condition or other implementation factors. UE just need to reports the applicable functionalities from UE point of view. |
| Futurewei | Yes. |  |
| NEC | Yes |  |
| vivo | No | Agree with Xiaomi to remove “final” as “~~final~~ applicable functionalities” |
| Apple | Partially Yes | Besides UE-side and NW-side additional condition, the model availability in device (i.e. the UE completes model training and downloading) also need to be met. Otherwise, the NW can’t know whether the functionalities are applicable and when they can be activated.  Thus, we suggest below change on Rapporteur’s proposal:  UE reports final applicable functionalities (applicable based on **~~both~~ model availability in device,** UE **side additional condition** and NW side additional condition) in Step 4, as a response to Step 3 |
| Huawei, HiSilicon | Yes with comments | Agree with Xiaomi to remove the wording "final" in the last step. |
| ZTE |  | See answer in Q2-1 |
| Mediatek | Agree with Apple | We agree with the Rapporteur's view that the UE should determine the final applicability of the AI/ML functionality, provided that the network-side condition, e.g., represented by the associated ID, is supplied alongside the inference configuration. This approach minimizes the need for additional handshaking, signaling, and potential latency that would result if the network were to make the final decision. |
| Lenovo | See comment | As commented in Q2-2, it seems similar approach as Option 1 in Q1-2. In Option 1 of Q1-2, it is upon UE to do filtering and NW makes the final determination. Here, it is upon NW to do filtering and UE makes the final determination. Not sure if we need different approaches supported, prefer to have one principle solution, even though we may have different variation in terms of proactive and reactive (if agreed). |
| Ericsson | Yes, but “final” not needed | Suggested changes to the figure:   * In the RRC Reconfiguration signalling just include “RRCReconfiguration (inference configurations including NW-side additional conditions, AIML functionalities for applicability reporting)”…..i.e. remove “NW-considered applicable”, because the NW cannot know at this stage whether a function is applicable or not. * First box: add “based on the received RRCReconfiguration” * We cannot preclude the case that the “applicability functionality reporting” can also include here other inference configurations/NW-side additional conditions, e.g if none of the inference configurations are ok, or simply if the gNB wants to use the RRCReconfiguration to inquire the UE about its applicabilities and needed inference configurations/NW-side additional conditions |

In below sections, let’s have some initial discussion on candidate RRC signalings used for Step 3-4. According to the contributions submitted RAN2#126 meeting. There are mainly two options proposed by companies:

**Option 1:** RRCReconfiguration/ RRCReconfigurationComplete (including RRCResume /RRCResumeComplete, etc) [1][9][15][18][21]

In this option, *RRCReconfigurationComplete* message is directly used to report the applicable functionalities in response to *RRCReconfiguration* in Step 3. Furthermore, [9][18][21] also mentioned that complete messages can be used when synchronization is lost between UE and NW, RRC state transition, etc.

**Option 2:** UAI (i.e. same as Approach #1 (proactive reporting), OtherConfig in RRCReconfiguration and UAI) [2][5][10][19][25]

In this option, UAI message can also be used in response to configurations received by UE in Step 3, indicating the applicable functionalities, i.e. reuse the same signaling as proactive reporting. By using UAI framework for reactive reporting, it also means network also needs at least configure the UE whether it is allowed to do applicable functionality reporting in Step 3.

#### Q2-4. Which option do you agree as signalings for Step 3 and Step 4?

**Option 1: RRCReconfiguration/ RRCReconfigurationComplete (including RRCResume /RRCResumeComplete, etc)**

**Option 2: UAI (i.e. same as Approach #1 (proactive reporting), OtherConfig in RRCReconfiguration and UAI)**

**Option 3: Others (please list the signaling message)**

|  |  |  |
| --- | --- | --- |
| **Company** | **Option 1/2/3** | **Comment** |
| OPPO | Option1 | See answer in Q2-1, only Option1 is applicable for reactive method. |
| Xiaomi | 1/2 | No strong view. Both options are feasible. |
| Futurewei | 2 | Option 2 can work for both proactive and reactive cases. Option 1 works too. But we have agreed to use Option 2 for the proactive case. We should use the same approach for both cases. |
| NEC | - | Option 1: On the top of reporting UE supported NW-side additional conditions via UE capability signalling and proactive applicable functionality reporting, we don’t see motivations to introduce this reactive reporting via RRCReconfiguration/ RRCReconfigurationComplete (including RRCResume /RRCResumeComplete, etc).  Option 2, we are OK with Option 2, however, it is the same as proactive reporting.  See answer in Q0-2, we think it would be good to use UE capability signalling to reduce AI/ML configuration latency and configuration overhead. |
| vivo | Option2/3 | Prefer to have unified procedure for reactive and proactive reporting, specifically, A UE have applicable functionalities may initiate the UAI procedure if it was configured to do so, upon it was configured to provide applicable functionalities, or upon change of applicable functionalities.  The former case (initial report) can be regarded as reactive reporting and the latter ones (upon change) can be regarded as proactive reporting.  We are also OK for the separate procedures. But we do not think Option 1, is appropriate for applicable functionality reporting. *UEInformationRequest /response* can be considered, which is used by the network to request the UE to report information. |
| Apple | Option 1 | 1. As we responded in Q2-1, in reactive reporting, associated ID and corresponding inference config are provided to the UE together in same *RRCReconfiguraiton* message, and the UE can immediately perform inference for applicable functionality. Thus, the UE can naturally include which functionalities are applicable in the response RRC message (i.e. *RRCReconfiguraitonComplete).* **It is unnecessary and redundant to use another RRC message (e.g. UAI) to report them as the UE already performs inference.** 2. We don’t think option 2 (i.e. UAI) can work for reactive reporting of neighbour cells. Existing UAI framework can only report UE’s **condition/information changes** **on serving cell, but it can’t work for reactive reporting of neighbour cells.** For example, in L3 measurement prediction for neighbour cells, the existing UAI framework can’t report applicable functionalities corresponding to the NW-side additional conditions of neighbour cells. So, if option 2 is adopted, we need to enhance existing UAI framework. 3. In Handover, option 2 (i.e. UAI) also can’t work for reactive reporting. In legacy, target cell configuration is included in HO command and the UE response *RRCReconfiguraitonComplete* message to target cell. It is straight forward that associated ID and inference configuration of target cell are also included in HO command, and the UE includes applicable functionalities in the response *RRCReconfiguraitonComplete* message towards target cell. As UAI message can’t be sent directly to target cell, it means existing UAI framework can’t work for handover case.   If RAN2 can extend existing UAI framework to support above neighbour cell and Handover case, we are also fine. But we assume that it will introduce significant spec change as it changes the fundamental assumption of UAI framework and put a new UE requirement (i.e. the UE needs to monitor condition/info change of both serving cell and neighbour cells). |
| Huawei, HiSilicon | Option 2 | We prefer to have a unified solution for both proactive and reactive approaches as it is sufficient and has less specification impacts. For Option 1, we wonder about the necessity.  The applicable functionality reporting means the applicability status may change dynamically, and then Option 1 may be inefficient in some cases. For example, at T0, the UE uses Option 1 for reporting and then the NW immediately knows that the UE has no applicable functionalities, but just after a short time, the UE has applicable functionalities (and vice versa). In this case, the NW does not know the change, and it will lead to inefficient signalling procedure between UE and NW.  In order to solve this issue, the NW may initiate multiple RRCReconfiguration procedures, and the UE could be able to send applicable functionalities via complete messages correspondingly. However, this may cause significant signalling overhead and NW resources. |
| ZTE | Either Option 1 or Option 2 ,not both | See our answer in Q2-1.  If the associated Id is adopt, we think the one unified solution is enough no matter the applicability reporting is reactive or proactive. |
| Mediatek | Option 1/2 | Even for proactive reporting, we think option 1 is possible. For example, the NW provides the associated ID in the *RRCReconfiguration* message, and UE can respond directly in the *RRCReconfigurationComplete* message whether/which AI/ML functionality is applicable.  But we are open to consider both option 1 and 2 for reactive reporting. |
| Lenovo | Option 2 | We tend to believe UAI based solution would be sufficient. The UAI can be triggered by UE anyway in case of any change of NW side additional condition e.g., via *RRCReconfiguration*. The motivation of using Option 1 *RRCReoconfigurationComplete* message seems be because of the latency consideration, but not sure if it is really necessary. |
| Ericsson | Option 1 | The reactive approach is used whenever we want the UE to react to specific inference configurations. It does not make sense to use UAI for it. The proactive approach has to be used mainly for the cases in which the UE wants to signal a change in its applicability conditions (due to UE conditions, or other received NW configurations). One benefit of reactive approach is that the UE can consider the applicable AIML configuration active straight away, without further NW configurations. If UAI is only UAI this cannot be achieved, as also Apple pointed out. However option 2, which is the proactive approach, should be integrated with the option 1, e.g. the network provides an inference configuration reactive approach, i.e. via classical RRCReconfiguration signalling (step 3), the UE applies it and starts doing AIML (step 4). Then at a later point in time the UE can signal changes in its applicability, e.g. the applied inference configuration is not applicable anymore. |

#### Q2-5. Please also list any missing information that you think needs to be transferred to UE or reported to NW in Step 3/4. If none, please skip this question.

|  |  |
| --- | --- |
| **Company** | **Comment** |
| Apple | * + - 1. As we responded in Q2-1, NW-sided additional conditions are always provided to the UE in the form of associated IDs, irrespective of proactive reporting or reactive reporting. Thus, we think **associated ID** are needed to be provided by NW.       2. Since inference configuration is already provided before applicable functionality reporting, we think it is sufficient for the UE to report whether they are applicable or not (e.g. the associated IDs whose corresponding inference configuration are applicable). |
| Ericsson | In step-4 we should not preclude the UE to report to the NW, the associated IDs/inference configurations, even if these associated IDs/inference configurations are not included in step-3. The step-3 can just be used by the gNB to require the UE to transmit its applicable functionalities and the related associated IDs, and then the gNB can use step-5 to configure, if interested, the needed inference configuration.  So the step 4 could:   * Include the applicable functionality/ies and the inference configurations/NW-side additional conditions selected by the UE for the inference (among the inference configurations/NW-side additional conditions included in the step 3). * Include other possible inference configurations/NW-side additional conditions that the gNB can configure to the UE (not included in the step 3).   Related to Apple comment 1: We agree that the NW-side additional conditions are part of the possible inference configurations in the reactive approach that the NW can provide to the UE. However, RAN2 does not need to discuss the details of this inference configuration. What to include there should be decided by RAN1. |
|  |  |
|  |  |
|  |  |
|  |  |

After Step 4 and determination of applicable functionalities, it is not clear whether the network needs to provide an update of network configurations or not.

The update of network configuration may not be needed, considering the UE itself knows the applicable functionalities and can directly use the network configuration in Step 3 without further update from network. On the other hand, the network may not have provided the full AIML resource configuration in Step 3 and may need to provide an updated configuration to UE according to the final applicable functionalities reported by the UE.

#### Q2-6. Do you think Step 5 is needed in reactive reporting? if yes, please comment on what is the difference with Step 3.

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO |  | See answer in Q2-1 |
| Xiaomi | Up to NW | NW may send *RRCReconfiguration* to update the NW configuration or other configuration. We don’t need to restrict NW implementation. |
| Futurewei | No | Does not seem necessary; does it provide any new information that the UE does not know? And how the UE is going to use it? |
| NEC | Yes | Step3 only configures NW-side additional condition related configurations (e.g., AI/ML beam resource configuration of Set A and Set B) whereas Step 5 should provide a full configuration for the activated functionality, for example, inference and/or performance monitoring related configurations, etc. |
| vivo | Yes | Especially for the type 2 conditions.  - Type 2: Set A/Set B related info, e.g., the beam index of set B. This information should be explicitly sent to UE for model inference. |
| Apple | Up to NW | Same view as Xiaomi, it is legacy behaviour that it is up to NW implementation whether/when to send another *RRCReconfiguration* to update the NW configuration or other configuration. It doesn’t make sense to restrict NW implementation. |
| Huawei, HiSilicon | Yes | For the difference between Step 3 and Step 5, our understanding is as below:  Step 3 is ‘otherConfig’ for the UAI containing additional configuration such as the assocaitedID and functionalities;  Step 5 contains the inference configuration that the network chooses for the UE inference operation. |
| ZTE |  | See answer in Q2-1 |
| Mediatek | Up to NW |  |
| Lenovo | Yes? | The inference related configuration should be provided to UE before the activation of AIML functionality.  From another aspect, it would be reasonable if NW provides functionality-specific configuration for inference after an AIML model being available at UE. Otherwise, it may be a waste of effort, as the AIML model may be never available.  If companies believe the availability is indicated by the applicability report, then the answer would be yes.  The same issue applies to proactive reporting scenario. |
| Ericsson | Up to NW | The step-5 can be used if for example the UE indicates in step-4 that the UE needs a different inference configuration (NW-side additional conditions) to make an AIML functionality applicable. Then the gNB can use step-5 can be used by the gNB to configure, if interested, the UE with the needed inference configuration. |

During RAN2 #126 meeting, it was agreed that, for a functionality to be applicable, it should have at least one model available within it. It is straightforward to conclude that applicable functionalities reported in Step 4 has an available model. During phase 1 discussion, companies have different understanding on whether to define available functionalities separately from applicable functionalities. Then, the question is whether a UE can report its non-applicable functionalities but with an available model via the same message used for applicable functionality reporting or not.

#### Q2-7. For reactive reporting, do you think UE can also include non-applicable functionalities with an available model in Step 4?

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO | Yes if reactive method definition is based on our understanding in Q2-1 | See answer in Q2-1 |
| Xiaomi | Comment | In current procedure, UE reports the applicable functionality or non-applicable functionality. Does rapp intend to introduce a new type of functionality? We don’t see the need to introduce more types of non-applicable functionality. |
| Futurewei | Comment | We don’t quite understand the question; why we want to report a non-applicable functionality with an available model? |
| NEC | No | Both proactive and reactive reporting are limited to applicable functionalities. |
| vivo | no |  |
| Apple | No | For both proactive and reactive reporting, we think the UE determines applicable functionaries **when all below conditions are met**:   * NW-side additional condition (i.e. the UE detects inference config and training config with same associated ID). * UE-side additional condition (e.g. current left memory/battery resource is sufficient to do inference). * Model is available in device.   It seems this question is whether we need to introduce another type of functionality which only meet above condition 3). If the understanding is correct, we think it is not necessary: at least 2) is up to UE implementation. Then, with this new functionality reporting, the NW can’t derive whether condition 2) is met in the UE side, and thereby can’t determine applicable functionalities. So, it is not useful. |
| Huawei, HiSilicon | See comments | If the question is on whether the UE should report the functionalities for which it does not have a model, then we should further understand whether there is anything the NW can do about it, e.g. configure model training etc. |
| ZTE | Too early to discuss | Too early to discuss, it is not sure why we need to take such information. |
| Mediatek | No | For the determination and reporting of AI/ML functionality applicability, the UE already considers the availability of the AI/ML model when determining applicability. Therefore, the necessity of additionally indicating the availability remains unclear. |
| Lenovo | See comment | Does it mean only applicable/non-applicable info of available models will be reported to NW, and for those not available, UE will not indicate anything? It seems to be a way of implying unavailability. |
| Ericsson | Yes, but question not clear | In all the options discussed above (proactive/reactive) one of the main objectives is for the UE to inform the network about the configurations that are needed to make an available AIML functionality applicable, so that the gNB can provide the necessary configuration to make such AIML functionality applicable. If the AIML functionality is already applicable, given the current configuration, then the UE just needs to report that the AIML functionality is applicable without signalling further info, e.g NW-side additional conditions, and the NW can just activate it. So the question in its current formulation appears unclear given the discussion on the previous sections about all the options for proactive/reactive.  The question should be instead what is the UE behaviour related to step-4 when the AIML functionality requested by the NW in step-3 is not available/trained. Should the UE use the step-4 to ask for training? Should the UE leave the applicable functionality reporting empty for such AIML functionality? Should the UE indicate that the AIML model is not available? We suggest discussing this issue. |

#### Q2-8. Any other comment on the above signaling flow of reactive reporting?

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

## Others

proactive vs. reactive

After discussing the above signaling, hope companies now have a clear understanding of what is proactive and reactive reporting. Based on above assumptions, the key differences between proactive and reactive reporting can be summarized as below:

|  |  |  |
| --- | --- | --- |
|  | **What is the trigger applicable functionality reporting?** | **Sequence between configuration and applicable functionality reporting** |
| **Proactive reporting** | Upon a change of condition(s) | Applicable functionality reporting -> Configuration |
| **Reactive reporting** | As a response to network configuration | Configuration -> Applicable functionality reporting |

#### Q3-1. Do you agree the key differences between proactive and reactive reporting are:

1. **trigger of applicable functionality reporting: upon a change of condition(s) (proactive) or as a response to network configuration (reactive)**
2. **whether the configuration of functionalities is provided after applicable functionality reporting (proactive) or before (reactive)**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO | No | See answer in Q2-1 |
| Xiaomi | Comment on the proactive trigger | We understand the trigger of proactive reporting upon a change of functionality applicability, rather than the condition. Because the condition change may not necessarily result in functionality applicability change. Upon condition change, UE shall first determine whether functionality applicability changes. If yes, trigger proactive report. If no, no need to trigger proactive report.  We are fine with other parts. |
| NEC | Partially Yes | For reactive reporting, sequence should be  NW-side additional condition related configuration -> Applicable functionality reporting -> full configuration |
| vivo | Yes for 1,  No for 2 |  |
| Apple | Yes with comments | We agree with the intention of Rapporteur, but we think “**the configuration of functionalities”** in 2nd bullet is not clear. To avoid misunderstanding, we suggest to make it clear that it is inference configuration. For example, we provide below suggested wording:   1. **trigger of applicable functionality reporting: upon a change of condition(s) (proactive) or as a response to network inference configuration (reactive)** 2. **whether the inference configuration of functionalities is provided after applicable functionality reporting (proactive) or before (reactive)** |
| Huawei, HiSilicon | No | In our opinion, there are two differences:  (1) the reponse messages are different  (2) even if the response messages can be the same, the usages are different. In reactive reporting, the UE reports the functionalities upon being configured with applicable functionalities reporting. While in proactive reporting, the UE additionally reports the applicable functionalities when the applicable functionalities change. |
| ZTE | No for 1  Yes for 2 | For triggering proactive reporting :  In the example of proactive reporting, it is straight forward that UAI (applicability reporting) is triggered by otherconfig of RRCReconfiguraion which does not match what the table says triggered upon condition change.  For triggering reactive reporting：  We do not agree the blindly configuring of the functionality is a rational implementation at NW side. |
| Mediatek | Yes with comment | Just as commented in Q2-1, the difference between proactive and reactive reporting lies in the timing of coordinating AI/ML functionality applicability relative to the inference configuration.   * In proactive reporting, the network provides the inference configuration to the UE after the coordination of the AI/ML functionality applicability. * In reactive reporting, the network provides the inference configuration to the UE before the coordination of the AI/ML functionality applicability.   For proactive reporting, one possible trigger is the availability of a new AI/ML model for a functionality. |
| Lenovo | See comment | Not sure if 1) and 2) are combined.  The fundamental difference of proactive and reactive would be if it is   * UE report triggered by UE (proactive) * UE report triggered by NW based on NW request/configuration (reactive)   In the case of UE determines the functionality applicability, in both proactive and reactive cases, UE needs to know enough info (NW configuration, additional condition) to determine the applicability of a functionality.  In addition, we believe the UAI based framework can serve the purpose of reporting the applicability, which can be considered as proactive/reactive depending on the interpretation. |
| Ericsson | 1. Yes 2. Changes needed | 1. Those are the fundamental scopes of the two approaches, which as mentioned above should coexist. The proactive is used for the UE to signal any change in the applicability (due to a number of reasons such as UE-side conditions), whereas the reactive is for the gNB to quickly/efficiently configure AIML 2. That is a direct consequence of 1). However, it cannot be precluded that also for the reactive approach there is a further reconfiguration after the applicability reporting (i.e. step-5 in the reactive reporting). The UE may be allowed by the network to signal in the reactive approach its recommended configurations/NW-side additional conditions, and then the gNB can configure it in a following message. So we suggest rewording 2) as follows: **the configuration of functionalities is provided after applicable functionality reporting (proactive) or before (reactive) and optionally after (reactive)** |

Initial Activation/Deactivation

During RAN2 #126 meeting discussion, following agreements are captured on applicable functionalities and activated functionalities:

|  |
| --- |
| RAN2 will support functionality activation/deactivation after inference configuration. FFS initial state of configuration and how activation/deactivation is achieved. |

After applicable functionality reporting, it is not clear what is the initial state (active/deactive) of the functionality after Step 5, and how an applicable functionality becomes an activated functionality. There are three options:

**Option 1**: The applicable functionality is activated by receiving configuration for applicable functionalities in Step 5 (if needed). If configuration is not provided by the network, it means the functionality is not activated.

**Option 2**: The applicable functionality is automatically activated if it is included in applicable functionality reporting (assuming the network configuration received in Step 3 is directly applied and the functionality is activated).

**Option 3:** A functionality is activated based on a field in RRCReconfiguration in Step 3 or Step 5 (indicating the functionality activation status), and additionally via L1/L2 based activation/deactivation signaling. L1/L2 based activation/deactivation signaling is up to RAN1.

#### Q3-2. Which option do you agree that an applicable functionality initially becomes an activated functionality after receiving configuration in Step 5?

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Proactive reporting:**  **Option 1/2/3** | **Reactive reporting:**  **Option 1/2/3** | **Comment** |
| OPPO | Option1/3 | Option1/3 | Option1 should be the baseline. Option3 can be considered for model switching/deactivation/activation. |
| Xiaomi | 1 | 1 | Option 1 is used for initial activation. Option 3 is used for subsequent activation/deactivation.  Although the same RRC message may be used in step 3 and 5, e.g. *RRCReconfiguration*, different IEs would be used to indicate the functionality in step 3 and 5. In step 3, the purpose is not to activate the functionality. |
| NEC | Option 3 | Option 3 | Should be in Step 5. |
| vivo | 1 or 3 | 1 or 3 |  |
| Apple | Option 1 as baseline | Option 1 as baseline | Option 1 obviously can work.  Whether Option 3 is needed depends on whether RAN1 agree to introduce L1/L2 activation signalling, which further depends on whether RAN1 identify requirement for dynamic activation, deactivation and switching. As usual, it is RAN1 decision and RAN2 can just wait for RAN1 conclusion. |
| Huawei, HiSilicon | Option 1 | Option 1 | Option 1 is reasonable. |
| ZTE | Waiting for RAN1 | Waiting for RAN1 | In our understanding, which option to go is depending on RRC signalling structure and granularity of functionality. At the current stage, RAN2 is not clear about either the detail RRC signalling structure (ASN.1) or the granularity of functionality, which needs to be firstly designed by RAN1. In this sense, we suggest to postpone the discussion until sufficient progress is made by RAN1. |
| Mediatek | Option 1 | Option 1 | Option 1 is the baseline. |
| Lenovo | 3/1 | 3/1 | RAN2 agreed before the following  1 For UE-sided model, for the functionality management, the “network decision, network-initiated” AI/ML management is supported as a baseline. The following can be considered further “UE autonomous, decision reported to the network”, “Network decision, UE-initiated” (i.e. proactive approach).  Since the applicability of AIML functionality could vary depending on the UE/NW side condition, performance etc. It would be good to support dynamic activation/deactivation of AIML functionality as in Option 3 or Option 1.  Besides, we don’t think L2 MAC CE based activation/deactivation solution would depend on RAN1. |
| Ericsson | Option 1 | Option 1 (baseline)  2 and 3 (possible for more efficiency) | Proactive: The legacy framework of UAI is used by the UE to signal its recommendation/preference to the gNB at any point in time, and then the gNB can decide whether to accept this recommendation and configure the UE accordingly. We need to follow here the same mindset, the UE cannot activate a functionality without previously informing the NW, since the NW does not know the point in time in which the UAI will be transmitted, and it is not guaranteed that the gNB can now accept the UE recommendation/preference.  Hence, only option 1 should be considered for proactive approach, followed potentially by explicit L1/L2 signalling for the actual activation.  Reactive: Unlike the proactive approach, here the gNB can provide the candidate inference configurations/NW-side addition conditions already in step-3. Hence in theory, the UE can just select one of these configurations, report the selected configurations to the gNB, and apply it straight away, without the step-5. So option 2 can also be adopted to avoid extra signalling. In our understanding option 3 is a way to configure the UE to adopt option 2 or option 1 (with the only difference that in option 1 RRC is needed, whereas option 3 proposes L1/L2). Is that correct understanding? |

## Positioning

During RAN2 #126 meeting online discussion, following agreements on positioning Case 1 were made:

1. The LPP Capability Transfer procedures (RequestCapabilities/ProvideCapabilities messages) are used to indicate supported AI/ML positioning capabilities. FFS how to handle dynamic capabilities, depending on further RAN1 progress and understanding of the functionality.
2. wait for RAN1 for associate ID discussion

Considering the exact signaling needs further RAN1 inputs, rapporteur suggests to only focus on the **general principle** for positioning during this discussion, e.g. exact LPP signaling (whether existing signaling or new enhancement) and NW-side/UE-side additional condition will **not** be discussed in this email discussion.

It was agreed in RAN2 #125bis meeting that proactive reporting and reactive reporting are applicable for both beam management and positioning use cases.

**Agreements for positioning and beam management**

1. Support proactive reporting of UE-sided applicable functionality, e.g., the UE reports its applicable AI/ML functionalities via UAI message/LPP message.
2. Support reactive reporting of UE-sided applicable functionality. The NW configures AI/ML functionalities via RRC/LPP message. FFS what the configuration contains. FFS how to report applicable functionality and what is applicable functionality

Based on the discussion of beam management, it seems there are some commonalities between positioning and beam management in terms of proactive/reactive reporting. Following assumptions for beam managements are also applicable for positioning, except the protocol is LPP between UE and LMF:

* + - 1. The key difference between proactive and reactive reporting is the trigger of applicable functionality reporting, and whether the configuration of functionalities is provided after applicable functionality reporting or before. This seems can also be applicable for positioning Case 1: i.e.

1) in **proactive reporting**, UE can report a change in applicable functionality/condition by sending applicable functionalities upon change via LPP signaling, network provides configurations of functionalities via LPP signaling **after** NW knowing applicable functionalities

2) in **reactive reporting**, applicable functionality reporting is provided as **a response to** the network configuration, where network provides configurations of functionalities via LPP signaling **before** NW receiving applicable functionalities.

* + - 1. Depends on Q2-7 if agreed, UE also sends functionalities with available model, but not applicable via LPP signaling in the same message used for applicable functionality reporting.
      2. Depends on Q3-2, the configuration after NW receiving applicable functionality reporting may or may not means the functionality is activated.

#### Q4-1. For positioning Case 1, do you agree with the above assumptions?

NOTE: Exact LPP signaling (whether existing signaling or new enhancement) and NW-side/UE-side additional condition will **not** be discussed in this email discussion.

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| OPPO | Yes for assumption 2 and 3 | For assumption 1, we have different view as commented in Q2-1 |
| Xiaomi | Comments for assumption 1  Yes for assumption 2 and 3 | We have the following understandings regarding the functionality reporting for positioning:   * The UE reports the applicable functionality, not the condition. * Proactive reporting implies that the reporting occurs without a network request, whereas reactive reporting is based on a network request. * The LPP is used between the UE and the LMF, we suggest using ‘LMF’ instead of ‘network’.   Therefore, we make some changes as below:  1) in **proactive reporting**, UE can report a change in applicable functionality/condition by sending applicable functionalities upon change via LPP signaling without LMF request, LMF network provides configurations of functionalities via LPP signaling **after** LMF NW knowing applicable functionalities  2) in **reactive reporting**, applicable functionality reporting is provided as **a response to** the network configuration LMF request, where LMF network provides configurations of functionalities via LPP signaling **before** LMF NW receiving applicable functionalities. |
| vivo | Yes for 1 with comments.  No for 2  Yes for 3 | For 1) and 2), We think the description can be simplified as:  1) in **proactive reporting**, UE can report a change in applicable functionality~~/condition~~ by sending applicable functionalities upon change via LPP signaling, ~~network provides configurations of functionalities via LPP signaling~~ **~~after~~** ~~NW knowing applicable functionalities~~  2) in **reactive reporting**, applicable functionality reporting is provided as **a response to** the network request ~~configuration, where network provides configurations of functionalities via LPP signaling~~ **~~before~~** ~~NW receiving applicable functionalities~~. |
| Apple | Comments to assumption 1,  Yes for assumption 2 and 3 | We agree with Xiaomi that “the UE reports the applicable functionality, not the condition”. The other part of assumption 1 is aligned with our understanding. Thus, we suggest below change:   * + - 1. The key difference between proactive and reactive reporting is the trigger of applicable functionality reporting, and whether the **inference** configuration of functionalities is provided after applicable functionality reporting or before. This seems can also be applicable for positioning Case 1: i.e.   1) in **proactive reporting**, UE can report a change in applicable functionality**~~/condition~~** by sending applicable functionalities upon change via LPP signaling, network provides **inference** configurations of functionalities via LPP signaling **after** NW knowing applicable functionalities  2) in **reactive reporting**, applicable functionality reporting is provided as **a response to** the network **inference** configuration, where network provides **inference** configurations of functionalities via LPP signaling **before** NW receiving applicable functionalities. |
| Huawei, HiSilicon | Yes for 2, 3  No for 1 | Xiaomi's revision can be considered.  Based on Xiaomi's revision, **we think 1) should remove "without LMF request"**, because as we discussed in section 2.1, the NW may still need to configure something to the UE. With the wording "without LMF request", the UE is freely to send the applicable functionality to LMF, which is out of LMF control and thus may lead to some negative imapcts to NW side. |
| ZTE | Comments for assumption 1  Comments for assumption 2/3 | To assumption 1,we think in AI positioning, reactive reporting is UE to report the capability in a solicited way (see TS37.355 section 5.1.1), and proactive reporting is UE to report the capability in a unsolicited way (see TS37.355 section 5.1.2). The spec will have no hard requirement on the time order of capability reporting(proactive or reactive) and configuration provision.  Figure 5.1.1-1: LPP Capability Transfer procedure  Figure 5.1.2-1: LPP Capability Indication procedure  For assumption 2, we think UE should only report the functionality that the UE currently supported.  For assumption 3, from RAN2 signaling structure perspective, AI/ML positioning functionality (usecase) can be managed as positioning method, i.e., activating a positioning method means activating the corresponding AI/ML positioning functionalities (usecases). usually LMF gives activation on a positioning method after UE reports it supports the positioning method. So we think assumption 3 for AI pos should be ‘ the configuration after NW receiving applicable functionality reporting ~~may or may not~~ means the functionality is activated’ |
| Mediatek | Yes for 1  No for 2  Yes for 3 | For assumption 1, Apple’s revision looks better. |
| Lenovo | Yes for assumption 2 and 3 | For assumption 1, we don’t think “whether the configuration of functionalities is provided after applicable functionality reporting or before” is the key difference for proactive/reactive.  The fundamental difference of proactive and reactive would be if it is   * UE report triggered by UE (proactive) * UE report triggered by NW based on NW request/configuration (reactive)   In the case of UE determines the functionality applicability, in both proactive and reactive cases, UE needs to know enough info (NW configuration, additional condition) to determine the applicability of a functionality. |
| Ericsson | Assumption 1: Yes, but changes need  Assumption 2: Yes, but the question should also be on what happens when the AIML functionality is not available  Assumption 3: Details can be further discussed after Q3-2 is addressed | Assumption 1: Related to 1) in assumption 1, we agree with Xiaomi rewording, but we also agree with HW, that similar to the UE also the LPP signalling should be configured by the LMF first. Related to 2) we do not need to preclude the possibility for the UE to signal in response to a NW configuration, the NW-side additional conditions to make the AIML model applicable, and then the gNB can provide the needed configuration  So we suggest the following rewording related to assumption 1:  Therefore, we suggest following changes below:  1) in **proactive reporting**, UE can report, upon LMF configuration, a change in applicable functionality/condition by sending applicable functionalities upon change via applicability functionality reporting in LPP signaling without LMF request, LMF network provides configurations of functionalities via LPP signaling **after** LMF NW receiving the applicability functionality reporting. ~~knowing applicable functionalities~~  2) in **reactive reporting**, applicable functionality reporting is provided as **a response to** the network configuration LMF request, where LMF network provides configurations of functionalities via LPP signaling **before** LMF NW receiving applicable functionality~~ies~~ reporting, and optionally after LMF receiving the applicability functionality reporting.  Assumption 2: As our reply to Q2-7, it is obvious for use that the UE should use the LPP signalling to inform the LMF about the NW-side condition to make the available functionality applicable. If the AIML functionality is already applicable, no extra info should be reported by the UE, rather the UE can just signal that the AIML functionality is applicable straight away. The question should instead be on what happens if the AIML functionality requested in the reactive configuration is not available. What the UE should report in this case in LPP?  Assumption 3: It is not critical to address this question at the moment. BM assumption does not need to be necessarily mapped to positioning protocols. We can address this after discussing Q3-2 |

#### Q4-2. For positioning Case 1, any other information needs to be considered during proactive/reactive reporting?

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| Apple | Yes | As we responded in Q2-1, NW-sided additional conditions are always provided to the UE in the form of associated IDs, irrespective of proactive reporting or reactive reporting. Thus, we think **associated ID** are needed to be provided by NW in positioning case 1. |
| Ericsson | Comments | Agree with Apple comment above. The NW-side additional conditions are part of the possible inference configurations in the reactive approach that the NW can provide to the UE. However, RAN2 does not need to discuss the details of this inference configuration. What to include there should be decided by RAN1. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Conclusion

# Reference

[1] R2-2404151 LCM for UE-sided model for Beam Management use case OPPO

[2] R2-2404185 Beam management UE-sided model LCM signaling Intel Corporation

[3] R2-2404275 On LCM for UE-sided model for Beam Management use case Qualcomm Incorporated

[4] R2-2404341 Discussion on the LCM for UE-sided model for Beam Management Fujitsu

[5] R2-2404370 LCM for UE-sided model for Beam Management use case TCL

[6] R2-2404390 Discussion on LCM for UE-sided model for Beam Management vivo

[7] R2-2404503 LCM for UE-sided model for Beam Management use case Interdigital Inc.

[8] R2-2404599 Discussion on LCM for UE-sided model for BM Xiaomi

[9] R2-2404637 LCM procedure of UE-sided model for AI/ML based beam management Apple

[10] R2-2404691 Discussion on LCM for UE-sided model for Beam Management use case CATT

[11] R2-2404816 LCM for AIML based beam management with UE-sided model Lenovo

[12] R2-2404817 Discussion on applicability-related information and additional condition Lenovo

[13] R2-2404902 UE side model functionality definition Sony

[14] R2-2404933 Discussion on LCM for UE-sided model Spreadtrum Communications

[15] R2-2404941 Functionality based LCM related to UE-side models for BM Nokia

[16] R2-2404957 Functionality identification and applicability related report LG Electronics

[17] R2-2404958 Functionality management for UE-sided model LG Electronics

[18] R2-2405025 Discussion on LCM for UE-sided model for BM CMCC

[19] R2-2405073 Discussion on LCM for UE-sided model NEC

[20] R2-2405180 Functionality-based LCM for UE sided model Samsung

[21] R2-2405184 Considerations on LCM for UE side Model for AIML Based BM ZTE Corporation

[22] R2-2405246 LCM Framework for UE-sided model SHARP Corporation

[23] R2-2405266 LCM for UE-side models for beam management Ericsson discussion

[24] R2-2405274 Discussion on LCM for UE-Side Models Futurewei Technologies

[25] R2-2405337 Discussion on functionality based LCM for UE-sided model for BM Huawei, HiSilicon

[26] R2-2405536 Considerations for moving forward with Functionality-based LCM Kyocera

[27] R2-2405665 Discussion on applicable functionality NTT DOCOMO, INC.

# Annex – RAN2 agreements on UE-sided model LCM

RAN2 #125bis meeting

**Agreements**

1. Which AI/ML-enabled Features/FGs and functionalities are supported should be standardized. The details wait for RAN1’s progress. “supported” means that the UE is capable of supporting the functionality and doesn’t mean neccesarily that the UE has the model available. FFS what functionality refers to.

2. Supported AI/ML-enabled Features/FGs and supported functionalities are included in UE capability.

**Agreements for positioning and beam management**

1. Support proactive reporting of UE-sided applicable functionality, e.g., the UE reports its applicable AI/ML functionalities via UAI message/LPP message.
2. Support reactive reporting of UE-sided applicable functionality. The NW configures AI/ML functionalities via RRC/LPP message. FFS what the configuration contains. FFS how to report applicable functionality and what is applicable functionality

3 FFS how the two approaches will be specified and whether we can combine them into one procedure. FFS how to report applicable functionality, what is applicable functionality, how the UE determines which function is applicable or not (if it is needed)

**Agreements:**

1 For UE-sided model, for the functionality management, the “network decision, network-initiated” AI/ML management is supported as a baseline. The following can be considered further “UE autonomous, decision reported to the network”, “Network decision, UE-initiated” (i.e. proactive approach).

2 “UE-autonomous, UE’s decision is not reported to the network” is not considered for Rel-19

RAN2 #126 meeting

**Agreements**

1 RAN2 will support functionality activation/deactivation after inference configuration. FFS initial state of configuration and how activation/deactivation is achieved. FFS what Deactivation refers to: examples discussed: 1) fallback to legacy 2) switching, etc.

2 We will work offline on the definitions for functionality types and define what is availability.

3 The UE will indicate the gNB/LMF whether the AI/ML functionality is available/applicable. For a functionality to be applicable at least there should at least one model available within it. FFS other details on what is applicability/non-applicability.

4 For NW-side additional conditions, RAN2 assumes that RRC signaling from gNB to UE can be designed for consistency between inference and training. RAN2 will wait for RAN1 input for further details. FFS if the same applies to positioning

5 For BM use case, As a baseline the UE determines whether a functionality is applicable. Existing UAI framework is used at least for proactive reporting of applicable functionality. FFS reactive