3GPP TSG-RAN WG2 Meeting #126 R2-24xxxxx

Fukuoka, Japan, May 20th -24th, 2024

Agenda Item: 8.1.2.2

Source: Intel, Samsung

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

Document for: Discussion, 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

The deadline for providing comments for phase 1 is June 11, 2024 Tuesday at 21:00 UTC.

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

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| Apple | Peng Cheng | Pcheng24@apple.com |
| vivo | Boubacar Kimba D.A. | kimba@vivo.com |
| OPPO | Jiangsheng Fan | fanjiangsheng@oppo.com |
| Xiaomi | Xing Yang | Yangxing1@xiaomi.com |
| ZTE | Fei Dong | Dong.fei@zte.com.cn |
| Ericsson | Felipe Arraño Scharager | felipe.arrano.scharager@ericsson.com |
| NEC | Satoaki Hayashi | satoaki-hayashi@nec.com |
| Lenovo | Congchi Zhang | Zhangcc16@lenovo.com |
| CATT | Tangxun | tangxun@catt.cn |
| Mediatek | Yuanyuan Zhang | Yuany.zhang@mediatek.com |
| Kyocera | Mitsutaka Hata | mitsutaka.hata.gt@kyocera.jp |
| Huawei, HiSilicon | Jun Chen | jun.chen@huawei.com |
| Intel | Ziyi Li | ziyi.li@intel.com |
| LGE | Soo Kim | soo.kim@lge.com |
| Sharp | Rudraksh Shrivastava | shrivastavar@sharplabs.com |
| NTT DOCOMO | Koki Yamashita | Kouki.yamashita.dz@nttdocomo.com |
| Qualcomm | Rajeev Kumar | rkum@qti.qualcomm.com |

# Phase 1: definition of functionalities

In RAN2 #126 meeting, RAN2 discussed the following definition for functionality types and decided to have more discussion to identify the need of such definitions and whether further update is needed to clarify the definition [1].

|  |
| --- |
| Proposal 2: RAN2 agree the following definition for functionality types as a starting point.  - *Supported/identified functionalities:* this refers to functionalities that UE can indicate by using UE capabilities.  - *Configured functionalities:* this refers to functionalities that gNB can configure UE for model inference and performing measurements for training purposes?. Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration.  - *Applicable functionalities:* this refers to functionalities that the UE is ready to apply for model inference. It can be considered as candidates for functionality activation.  - *Activated functionalities:* this refers to functionalities that the UE starts predicting beam results via model inference. |

In this discussion, it would be good to discuss each functionality type.

## Supported functionalities

|  |
| --- |
| *Supported functionalities:* this refers to functionalities that UE can indicate by using UE capabilities. |

The moderator think that we can simplify the name from supported/identified functionalities to supported functionalities for convenience of discussion. Please comment if it is not ok.

**Q1: Do you agree that supported functionalities refer to functionalities that UE can indicate by using UE capability signaling and gNB/LMF can configure?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | Partial Yes | We prefer not to couple the “supported functionalities” with “configured functionalities” (i.e. the 2nd half sentence “and gNB/LMF can configure” can be removed):   1. As discussed in Q2/Q3, it is not clear whether the “gNB/LMF can configure” means “configure for training” and/or “configure for inference”. This may bring additional ambiguity. 2. Since UE capability is RAN2 expertise, we think RAN2 have common understand what “indicate by using UE capability signalling” means (i.e. UE supports the feature and NW can configure the feature). So, the 2nd half sentence is redundant.   Thus, we suggest below change:  **supported functionalities refer to functionalities that UE can indicate by using UE capability signaling ~~and gNB/LMF can configure~~** |
| vivo | Yes | The original one is OK and agree with Apple that “gNB/LMF can configure” is not needed as the gNB/LMF should configure based on applicability/availability rather than supported. |
| OPPO | Partial Yes | We understand supported functionalities means the functionalities indicated via UE capability signalling, i.e. those functionalities are already implemented and tested from product engineering perspective, but it does not necessarily mean that NW can configure UE only based on supported functionalities, because RAN2 already agreed the understanding that UE may or may not have the model available for the supported functionalities. If NW configures a UE supported functionality but UE does not have the model associated with this functionality when receiving the configuration, what will happen from UE side? Trigger RRC re-establishment or delay the functionality until making the model available? Neither way is desirable from system point of view.  We notice that this situation for AI functionality is different than legacy functionality, i.e. non-AI functionality, for non-AI functionality, once UE reports the supported functionality via UE capability signalling, NW can be sure to configure any UE supported functionality if NW supports the functionality also, but when it comes to AI functionality, the working logic and assumption is totally different as the model availability is the additional aspect NW has to consider before making any configuration, otherwise, there is a risk to trigger RRC re-establishment or delay the functionality until making the model available as mentioned above.  So we suggest the following simple definition for supported functionalities:  **Supported functionalities refer to functionalities indicated via UE capability signaling.** |
| Xiaomi | Yes | We understand supported functionalities are static, which would not change dynamically. We would support to keep the ‘gNB/LMF can configure’ part. We don’t see the use case for NW to configure a functionality which is not supported by UE. |
| ZTE | Partial Yes | We have some sympathies with Apple’s suggestion.  In RAN2#125bis meeting, the below agreements have been achieved for the supported functionality reported by UE capability:  ***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.***  In our understanding, if there is no model is available at UE side for the supported functionality reported in UE capability , there is no need for gNB/LMF to configure it. In this sense, we would like to remove the ‘gNB/LMF can configure’ since it implies that NW may configure the functionality without any available models to the UE which is a bad implementation. |
| Ericsson | Partially yes | Agree with Apple. The gNB/LMF configuration should rather be based on the UE applicability reporting, and on whether the model is available at the UE. Capability signalling should be instead used by the UE to indicate the functionalities that the UE is capable of, i.e., irrespective of whether the AIML functionality is applicable/available.  Suggested rephrasing: **"supported functionalities refer to functionalities that the UE is capable of, and can be indicated by using UE capability signaling”** |
| NEC | Partial Yes | Our understanding is that the gNB/LMF should only configure “applicable functionalities”. We also prefer to just say “Supported functionalities refer to functionalities indicated via UE capability signaling” (i.e., remove “and gNB/LMF can configure” part.) |
| Lenovo | Partial Yes | RAN2 had the following agreement before  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.  Besides, we believe for a supported functionality, it is reasonable to assume at least one model for this functionality is available at UE/UE-side already. We suppose RAN2 can have a definition based on the above agreement, e.g.   * **Supported functionalities refer to the functionalities indicated by UE capability signalling that UE is capable of and has at least one model available for the functionality.** |
| CATT | Partial Yes | We agree with Apple to remove “and gNB/LMF can configure” part, as in our view whether it can be configured depends on whether it’s applicable.  RAN2 agreed that “supported” means that the UE is capable of supporting the functionality and doesn’t mean necessarily that the UE has the model available, and for a functionality to be applicable at least there should at least one model available within it. |
| Mediatek | Partial Yes | Agree with Apple. The term 'supported functionality' does not necessarily suggest configurability. It also relates to the assumption of either proactive or reactive reporting for UE-side applicable functionality. Assuming proactive reporting, the gNB/LMF configures the AI/ML functionality following the UE capability and UE-side functionality report. In the case of reactive reporting, the gNB/LMF determines the AI/ML functionality based on the UE capability report alone. However, it needs to evaluate whether reactive reporting is necessary. |
| Kyocera | Partial Yes | We prefer to remove that "and gNB/LMF can configure", since we think it is not clear whether gNB/LMF can always configure functionalities that UE indicates in UE capability. |
| Huawei, HiSilicon | Partial Yes | OK with Apple's change.  For the wording "**and gNB/LMF can configure**", the meaning is unclear. This configuration can be everything. For example, after the UE reporting the applicable functionality, the NW configuration could mean that NW wants to enable one functionality. For monitoring, there may be some NW configurations for the UE for a functionality, but it is a separate topic.  **So we also suggest to remove "and gNB/LMF can configure".** |
| Intel | Yes | In general, the above statement is ok according to RAN2 #125 meeting agreement:   |  | | --- | | Supported AI/ML-enabled Features/FGs and supported functionalities are included in UE capability. |   However, there’s still some ambiguity whether the functionalities need to have an available model or not. However, this also depends on whether model transfer/delivery is needed or not. If model transfer/delivery is not needed, it is straightforward to assume the UE reports available models via UE capability, otherwise, functionalities reported via UE capability is supported functionalities but may not have available models, further configuration/procedure to support model transfer/delivery is needed.  Therefore, we think it would be better to remove “gNB/LMF can configure” for now before we understand more about how/when/what network configures to the UE and whether the model is assumed available at the UE side when reporting UE capabiltiy. |
| LGE | Partial Yes | Agree with Apple’s change. Supported functionality means that the UE is capable of a relevant functionality, but does not mean that the UE has a model and/or the appropriate configuration for that functionality. Otherwise, the capabilities may change dynamically depending on the presence of model or availability of configuration from the network. |
| Sharp | Partially Yes | Supported functionalities refer to functionalities indicated via UE capability signaling." Model availability and functionality applicability should be considered separately and should not be conflated.  Supported functionality doesn’t necessarily mean that the model is available. It is possible that the model is available but needs to be updated. Since the UE has limited space, it may not be able to maintain and store all the models associated with supported functionality. If the model is required, it may acquire the model via model transfer/delivery procedures. |
| DOCOMO | Partially Yes | Agree with Apple’s change. We think gNB/LMF configuration should be based on applicable functionality. |
| Qualcomm | Yes | *Configured functionalities:* this refers to functionalities that gNB has configured UE (ie current configuration).    *(In figure above: Applied-> applicable).*  RAN2 has a choice to allow the configuration to be only part of the applicable functionalities, or to allow the configuration to cover applicable and non-applicable functionalities.  On the other hand, only applicable functionalities can be activated.  While we agree that traditionally only those features are configured at the UE that can be used at the UE immediately. However, traditionally supported implied applicable. However, in the AI/ML based procedures, if network is allowed to configure functionalities that are applicable, then the configuration can be very limiting. This will also require frequent configuration from the network, which should be avoided.  The argument above that if model is not available at the UE, then it may result in RLF/reestablishment is not correct. Note that activation/deactivation/fallback/switching procedures are defined such that if the AI/ML models for functionalities are not available/applicable, then legacy procedures can be used. |

## Configured functionalities

|  |
| --- |
| *Configured functionalities:* this refers to functionalities that gNB/LMF configured to UE. UE for model inference and performing measurements for training purposes? Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration. |

As commented during RAN2 discussion, the moderator changed “gNB can configure” to “gNB configured”. And, LMF is added to cover positioning use case. Please comment if the change is not acceptable.

Please note that the second part (“Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration.”) will be discussed in Q5 in Section 2.3.

**Q2: Do you agree that configured functionalities refer to functionalities that gNB/LMF configured to UE?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | No | We doubt whether RAN2 really need this definition:   1. In 3GPP, we only specify UE behaviour under NW configuration. Thus, the definition suggested by moderator is obvious and doesn’t bring useful information. 2. As Q3 discussed, the configured functionalities may be understood as inference configuration or training configuration (and even performance monitoring configuration, although it was not mentioned in online discussion). So, the definition suggested by moderator may bring further ambiguity or further clarification. 3. We think at least stage 3 specification (either RRC or MAC) doesn’t need this definition because what NW configured to UE is always clearly specified in RRC and MAC spec. Thus, whether it is configuration for training or inference or monitoring should be crystal clear from text before and after in stage 3 specification. 4. In our understanding, the key controversial issue (or AI/ML specific issue) is the boundary among the following 3 terms: “supported functionality”, “applicable functionality” and “activated functionality”. Thus, it seems sufficient to define these 3 terms.   Thus, we think RAN2 don’t need definition of “configured functionalities”. |
| vivo | See comment | From our understanding, the term “configured functionality” is introduced for NW to configure the functionality in advance and activate it when needed.  For BM, the necessity of the definition can be FFS, depending on whether the configuration in advance is needed.  For positioning use case, the definition is not needed. We already agreed to reuse the existing LPP Location Information Transfer procedure (*RequestLocationInformation*/ *ProvideLocationInformation* messages) for providing the results of the UE-sided model inference operation, thus no configuration in advance is needed and LMF can directly request to activate the functionality. |
| OPPO | partially Yes | The suggested definition on configured functionalities does not make much sense as it’s too obvious if considering this terminology alone.  What really matters is the boundary among supported functionalities, available functionalities, configured functionalities, applicable functionalities and activated functionalities，because the boundary will decide how AI functionalities works in the whole picture, we understand the following relationship should be the baseline for our discussion:    For instance, UE supports 100 functionalities from UE capability signaling point of view, but currently 40 out of 100 functionalities, UE has the corresponding models available due to memory limitation; Then based on NW additional condition and UE inside additional condition, gNB knows 20 out of 40 functionalities is applicable based on UE reporting; then gNB configures 10 out of 20 functionalities to UE side and activate 5 out of 10 functionalities for this UE.  Based on above, we suggest the following definition for configured functionalities:  **configured functionalities refer to functionalities that gNB/LMF configured to UE, all configured functionalities shall be appliable and ready for activation from NW and UE perspective.** |
| Xiaomi | Yes with Comment | We understand configured functionalities can be used to trigger reactive applicable functionality report. However, if proactive applicable functionality report is used, NW can directly activate applicable functionality reported by UE. Configured functionality seems to be unnecessary in this case. We suggest to further clarify configured functionality is only used in reactive applicable functionality report. |
| ZTE | No | In our understanding, the configured functionalities is simply referring to the functionalities those have been configured to the UE. There is no need to over interpret this term since we have a lot of similar thing ( for example, configured TCI state, configured SCG, configured…., we should not always make a clear definition for those terms since we already have common understanding on what is configured). |
| Ericsson | No need to discuss this | Agree with other companies’ comments above, i.e., the intention of this question is not clear. A configured functionality is just a functionality that it is configured to the UE from an RRC point of view. However, this is obvious, and we do not need to discuss/agree to this. |
| NEC | See comments | Firstly, we think NW should only configure the applicable functionalities, i.e., configured functionality should be a subset of applicable functionalities. We understand that “available functionality” refers to supported functionalities matching the UE side additional conditions whereas “applicable functionality” refers to available functionality matching the NW side additional conditions. Based on this understanding, we agree with OPPO’s analyses on the relationship among “supported, available, applicable, configured, activated functionality”.  Then, we suggest following:  Configured functionalities: this refers to applicable functionalities that gNB/LMF configured to UE for model inference and performance monitoring. |
| Lenovo | See comment | We had the following agreement during the meeting:   * 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.   We understand “configured functionalities” means all the network configurations (e.g., SetA/SetB beam configuration, reporting configuration) needed to perform AIML inference are provided from the NW to UE. In other word, “configured functionalities” are ready to operate once determined to be applicable and then being activated. Maybe:   * **Configured functionalities means functionalities that are configured with the complete NW configurations needed to perform AIML inference, and they can be activated once determined to be applicable.** |
| CATT | No | We agree with Apple’s analysis. The applicable functionalities reported by UE represent all the functionalities that the current UE and NW can apply, e.g., based on the UE and NW side additional conditions. And the NW could directly activate one model among these models. Therefore we think this definition is unnecessary. |
| Mediatek | No | Agree with Apple, ZTE and Ericsson. The definition of this terminology is not needed. |
| Kyocera | See comment | We understand that “configured functionality” indicates a configuration from gNB/LMF to UE, and this is a quite usual practice. Therefore, we do not find a reason to discuss this definition. |
| Huawei, HiSilicon | See comments | In our opinion, the wording "**functionalities that gNB/LMF configured to UE**" is quite general. We can firstly clarify what is configuration, and what is the purpose of the configuration.  For example, if the NW would like UE to report all or some supported functionality for applicable functionality reporting, we may call them "configured functionalities". For other cases, there may be different usages.  **In summary, we suggest to clarify the content and the purpose of the configuration for the terminology "Configured functionalities".** |
| Moderator |  | I agree that the term of configured functionalities is very obvious because gNB will configure functionalities. In that sense, we may not need to specify it to the specification.  However, would it be good to have the term to have phase 2 discussion?  I already see the potential different view whether gNB can configure applicable functionalities only or not .  Based on Oppo’s figure (nice figure!), gNB configures only applicable functionalities. However, if we consider reactive approach, all the configured functionalities may not be applicable before gNB receives applicability related information/reported applicable functionalities from UE.  In addition, one of next discussion would be what is the relationship between support functionalities and what gNB can configure (which is configured functionalities).  Potential way forward: we can keep configured functionalities but explicitly note that the definition is not specified or FFS. |
| Intel | See comment | The definition proposed by rapporteur is technically correct, however, in our understanding, we still need to discuss based on what information the network can provide configuration (e.g. for training/inference/monitoring) to UE. After receiving the configuration from network, such functionality can be called as “configured functionality”, as in legacy. This is not a new concept for AI/ML functionalities. This could be straightforward and no need to have a separate definition.  It seems no need to specify this terminology. |
| Lenovo | Would be helpful to clarify whether it is “fully configured” or “partially configured” | As we commented earlier and also in Q5.  It would be good to clarify when we are saying “configured functionality”, whether   * It is partially configured, e.g., SetA/B beam configuration is provided to UE, so that UE could determine the applicability. But UE doesn’t have all required NW configuration for AIML inference in this case. * Or it is fully configured with all NW configuration needed to perform AIML inference once activated.   We understand Huawei also have similar question from their comment. |
| LGE | See Comment | The terms "proactive report" and "reactive report" might not be explicitly specified in the specifications, but they could be appropriately specified in relevant sections for each report.   * In the case of "reactive report", it can be specified as a report on applicable functionality to be configured/(de)activated among configured functionalities. * In the case of 'proactive report', it can be specified as a report to be configured/(de)activated for a certain functionality, i.e., configured functionalities may be a subset of applicable functionalities.   Therefore, it may be necessary to have a common understanding within RAN2 about the meaning of the configured functionality.  Regardless of "Reactive" or "Proactive" reports, the UE still needs to receive configuration for the activation (inference) of a functionality. Therefore, we support the definition proposed by rapp excluding model transfer related contents.  *Configured functionalities: this refers to functionalities that gNB/LMF configured to UE. UE for model inference. ~~and performing measurements for training purposes?~~ Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration.*  If the majority does not specify it, we can follow that. |
| Sharp | No | Agree with the companies above, no need to over complicate this term. Configured functionalities refer to features that have been fully set up with all necessary network configurations for AI/ML model inference and performance monitoring. It can be activated as soon as they are deemed applicable. |
| DOCOMO | See comment | It is unclear what timing of configuration is defined by “configured functionality”. If this “configured” means the configuration for functionality activation, “configured functionality” and “activated functionality” have the same meaning. Therefore, we think the discussion for this terminology is not needed.  In our understanding, there are some situation network configure to UE. So, if we discuss this terminology, more clarification is needed, e.g., configured functionality “for applicable functionality reporting” or “for functionality activation”. |
| Qualcomm | Yes | The answer is yes to Q2. We also agree with the analysis from Apple, ZTE, and Ericsson, no new definition needs to be added in stage 2 or stage 3, but this is the understanding of “Configured functionality”. |

**Q3: do you agree that this can be used for both UE-side model inference and training purpose?**

If it is NO, please indicate your preference e.g. only model inference for now or any suggestion for RAN2 progress.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple |  | See our comment to Q2. |
| vivo | No | The discussion of data collection for UE-sided model training is still ongoing and controversial. The motivation to involve model training in configured functionality is not clear.  Suggest focusing on the model inference for now and model training can be revisited when needed. |
| OPPO | No | Only for UE-side model inference for now as inference is usually our focus for configuration. We can further clarify training part when it’s clear enough for inference part. |
| Xiaomi | Only inference for now | For training, we are not sure whether functionality is needed. It’s possible the data collection is achieved by legacy measurement frame work, which is not related to functionality explicitly. |
| ZTE | No | we do not see the configured functionality can be used for model training purpose. For example, the data collection for model training need both set A and set B information for the spatial beam management, but the configured functionality only can provide the set B information, how UE can use set B only information to perform the model training.  So the configured functionality is just for inference, if UE really think only set B information can be used for model training, it is up to UE implementation and no need to capture in the definition of the term. |
| Ericsson | No | We sympathise with previous comments. The configuration for training and inference are two separate configurations, since they are intended for different LCM phases. We should not mix them together. |
| NEC | No | Model inference and performing measurements for training purposes are different phase, we also prefer to focus on model inference firstly. |
| Lenovo | No | Agree with OPPO. |
| CATT | No | Data collection from UE for model training may be aware by the UE or not, but UE-side model inference should be known and applicable by UE. So we share the view that configuration for training and inference are two separate configurations. And if the “configured functionality” needs to be defined, it is just for inference. |
| Mediatek | Only for inference | Agree with Ericsson that the configuration for training and inference are two separate configurations. |
| Kyocera | FFS | We think that including training is premature due to insufficient discussion. |
| Huawei, HiSilicon | No | It seems straightforward to focus on UE-side model inference for now.  For UE-side model training, it is a study objective and it is still under RAN2 evaluations. **So we suggest to not involve training for now.** |
| Intel | Yes | As we commented in Q2, same as legacy, as long as the UE receives any configuration from network (including training, inference and monitoring), it can be called as configured functionalities. Hence, we feel it is a general concept for both model inference and training. To clarify, the training configuration (e.g. training data collection configuration) may include (depending on the solution) how UE measures/collects data, how UE reports data, etc. But agree for training part, we may only consider solution 1b/2/3, since solution 1a is transparent to 3GPP. |
| LGE | No | The configuration for collecting training data and the configuration for inference are separate. Applicability depends on whether activation(inference) is possible, and in this sense, the configured functionality can only be considered in terms of inference. |
| Sharp | No | Share similar views as mentioned by other companies above. The configurations for training and inference should be kept distinct, as they pertain to different lifecycle management phases. They should not be combined. |
| DOCOMO | No | We think we should focus on only model inference for now. |
| Qualcomm | No | Agree with Ericsson.  Also,we should not be discussing a “training configuration” which implies a configuration about how training is performed. This is out of scope of this work item. |

## Applicable functionalities

|  |
| --- |
| *Applicable functionalities:* this refers to functionalities that the UE is ready to apply for model inference. It can be considered as candidates for functionality activation. |

It is also noted that RAN2 made a following agreement regarding applicable functionalities.

|  |
| --- |
| 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. |

**Q4: Do you agree that applicable functionalities refer to functionalities that the UE is ready to apply for model inference and they can be considered as candidates for functionality activation/deactivation?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | Yes with comments | We understand that “the UE is ready to apply for model inference” means the UE:   1. It has successfully completed model training compliant with UE’s internal status. And the trained model is already in UE device. 2. The dataset which was used for training the model is consistent with current inference configuration.    * Note that if it is not consistent, the UE can’t be regarded as “ready to apply” because current gNB is not aligned with UE and so can’t activate it.     Although we understand these are details which will be discussed in phase 2, it is better to confirm current definition can cover them. |
| vivo | Yes, with comments | The applicable functionalities are candidates for functionality configuration as well. In addition, deactivation is not needed as the deactivation may happen when the functionalities get inapplicable. Thus, suggest refining as:  **can be considered as candidates for functionality configuration (if needed) /activation.** |
| OPPO | partially Yes | Based on our picture raised in Q2, we understand applicable functionalities means these functionalities are ready to be configured and activated from both UE and NW point of view, but current definition only focuses on UE side, we think it’s not complete, so we have the following definition suggestion for applicable functionalities :  **applicable functionalities refer to functionalities that NW is ready to configure and the UE is ready to apply for model inference, and they can be considered as candidates for functionality activation/deactivation from both UE and NW point of view.** |
| Xiaomi | Yes | We agree with the principle proposed by rapp. How to determine the applicability can be up to UE. Because UE vendor may consider different UE implementations. |
| ZTE | Yes to have a definition, No for the current definition | We agree to have a definition of the applicable functionality. But the current definition is too complex to understand, we would like to make it more straightforward:  **Applicable functionalities refer to** **the functionalities those have available models and can be considered by UE and NW to be applicable for activation at the time beings.** |
| Ericsson | Yes, with modifications | We have to make clear that the functionality is applicable when it is available, and also under the condition that the inference configuration fits the training dataset. Hence, we propose clarifying as follows:  *Applicable functionalities:* refers to UE functionalities that the UE has available and that can be applied under a certain RRC configuration.  Similarly, we should define what is non-applicable functionalities:  *Non-applicable functionalities:* refers to UE functionalities that are not available in the UE and that cannot be applied given the current RRC configuration. |
| NEC | Yes with comment | Upon receiving the applicable functionality from the UE, the NW determines/sends the corresponding configuration for those candidates for activation. Therefore, suggest changing to:  Applicable functionalities: this refers to functionalities that the UE is ready to apply for model inference. It can be considered as candidates for functionality activation if configured by the NW. |
| Lenovo | Yes with comment | We believe the applicable functionalities are more about activation rather than deactivation. Also, we tend to believe when a functionality is determined to be applicable, it doesn’t necessarily mean the functionality is fully configured beforehand. For example, the SetA/B beam configuration can be provided to UE for applicability determination, while the CSI reporting configuration is not provided yet. Maybe:   * **Applicable functionalities refer to functionalities with a trained model that UE can apply for AIML inference under current condition(s)/configuration(s) and they can be activated once all configurations needed are provided by NW.** |
| CATT | Yes | Agree with Rapporteur. |
| Mediatek | Yes with comment | We think the applicable functionality needs to meet following requirements:   1. The functionalities have available models at the UE side 2. The functionalities are ready to be applied for inference and activated from UE perspective (considering the agreement that the UE will indicate the gNB/LMF whether the AI/ML functionality is available/applicable)   The definition may be revised as follow:  **Applicable functionalities refer to those for which AI/ML models are available at the UE and which the UE can consider for functionality activation.** |
| Kyocera | Yes | We agree with the definition suggested by the rapporteur in Q4. It makes it easier to manage which functionalities in NW will be activated or deactivated. |
| Huawei, HiSilicon | Yes with comments | The definition proposed by the rapporteur suggests that there are always two stages to make the UE start applying the inference for a functionality:   1. Functionality configuration 2. Functionality activation   In our understanding, this two-step approach has not been discussed in RAN2 before, and the meaning of "functionality configuration" is unclear (see our comments in Q2). We can start with simple definitions and details can be added when we know which solutions we apply.  In addition, "**functionality activation/deactivation**" may lead to some misunderstandings, e.g. it may be about using L2 signalling to let UE activate/deactivate a functionality, but RAN2 never discussed it before. For a specific feature, RAN2 normally use configuration/release of it at RRC signalling level, i.e. a feature is activated when it is configured, and a feature is deactivated when it is released.  **In summary, we suggest to modify the wording into:**  **Applicable functionalities refer to functionalities that the UE is ready to be configured and activated for model inference.** |
| Intel | Yes with comments | The applicable functionalities, in our understanding, means:  According to the UE-side additional condition and/or NW-side additional condition, the AI/ML functionality is ready to apply for model inference. Based on the applicable functionality information reported by the UE, NW can then provide configuration for inference.  However, we observe there’s some different understanding about when the configuration for inference is provided to the UE, for that part, we suggest to discuss in phase 2 based on signaling framework.  With that, we propose to consider below definition:  **applicable functionalities refer to functionalities that the UE is ready to apply for model inference ~~and they can be considered as candidates for functionality activation/deactivation~~** |
| LGE | Yes with comments | To align with Q2 answer, we sympathy with Huawei’s change.   * In the case of "reactive report", it can be specified as a report on applicable functionality to be configured/(de)activated among configured functionalities. * In the case of 'proactive report', it can be specified as a report to be configured/(de)activated for a certain functionality, i.e., configured functionalities may be a subset of applicable functionalities.   Additionally, since the possibility of model inference already implies that the UE has an available model, additional explanation about the existence of the model seems unnecessary. |
| Sharp | Yes, with comments | Applicable functionalities refer to UE capabilities that are available and can be applied or activated through RRC configuration. The UE may be able to determine the applicability of the functionality. The applicability may be determined based on UE and/or network side additional conditions, including also the UE side conditions such as e.g., based on Hardware resource availability, processing power etc.  As we mentioned in the Q1 response: Model availability and functionality applicability should be considered separately and should not be conflated. |
| DOCOMO | Yes with comment | In our understanding, applicable functionality means the UE has the appropriate model and is ready to apply for model inference. This means UE can start model inference immediately when NW configure to activate the functionality. And once the functionality becomes applicable, it will not become non-applicable generally, although it can become deactivated (or low-performance). |
| Qualcomm | Yes | Agree with Rapporteur. |

In relation to configured functionalities, configured functionalities may or may not be applicable immediately upon configuration depending on proactive/reactive approach. For example, in one way, UE already provides applicable functionalities/applicability related information and gNB can configure applicable functionalities. The other way would be that UE provide applicable functionalities/applicability related information after receiving configured functionalities from gNB and hence, configured functionalities may not be applicable immediately upon configured functionalities. We can remove “depending on proactive/reactive approach” in the definition as it will be discussed further in Phase 2 and there is no need to add as a part of definition.

**Q5: do you agree that configured functionalities may or may not be applicable immediately upon configuration?**

If it is NO (or partially Yes), please comment what relationship you envision between configured functionalities and applicable functionalities.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple |  | See our comment to Q2. We think RAN2 only need to define “**supported functionalities**”, “**applicable functionalities**” and “**activated functionalities**”, i.e. no need to define “**configured functionalities**”. |
| vivo | No | NW should configure AI functionalities based on the applicability/availability indication from UE. Otherwise, the configuration may be useless if the functionality is always unavailable. |
| OPPO | No | We think the following scenario is invalid and should be deprioritised:   * The other way would be that UE provide applicable functionalities/applicability related information after receiving configured functionalities from gNB and hence, configured functionalities may not be applicable immediately upon configured functionalities.   Why a NW would like to blindly configured potential non-applicable functionalities to UE before knowing which functionality is applicable from UE point of view. It’s more like a try-and-error logic, which should be avoided from 3GPP system perspective. More addition, what will happen from UE side when a functionality is not applicable when configured by NW? Trigger RRC re-establishment or delay the functionality until making the model available or something else? Neither way is desirable from system point of view.  Based on our picture raised in Q2, we understand configured functionalities shall be applicable from both UE and NW perspective.  The following scenario should be the baseline for our discussion:   * in one way, UE already provides applicable functionalities/applicability related information and gNB can configure applicable functionalities. |
| Xiaomi | Yes | We agree with rapp configured functionalities can be un-applicable when it’s used to trigger reactive applicable functionality report. |
| ZTE | No | We do not want to couple the configured functionality and applicable functionality. To me, the configured functionality is dependent on the RRC configuration, the applicable functionality depends on whether the functionality is applicable to the current UE and NW additional conditions. The former one is related to a time period of RRC configurations, the last one is related to a time period of scenario change. These are two different things, we should not couple them. |
| Ericsson | No need to discuss this | We agree that when functionalities are configured, these may (or not) be applicable when being configured. However, as pointed out by ZTE, we should not couple the applicability and the configuration together. The functionality configuration is an RRC procedure, and it is left to NW implementation, whereas the determination of the applicability is based on UE decision.  A configured functionality may or may not be applicable, but also a functionality not yet configured may or may not be applicable.  Hence, we see no need for RAN2 to discuss this. |
| NEC | No | NW should only configure the applicable functionalities, i.e., configured functionality should be a subset of applicable functionalities. |
| Lenovo | See comment | First, it could be helpful to clarify “configured functionality” means **fully configured** for AIML inference, or **some configuration** is provided to UE.  In our understanding, some of the functionality configurations would be needed to determine the applicability, e.g., SetA/B beam configuration. But that does not necessarily mean an applicable functionality must be fully configured beforehand.  Thus, we don’t see a strong linkage between when a functionality is “fully configured” and when a functionality is “applicable”. |
| CATT | No | If the “configured functionality” needs to be defined, we think NW should only configure the applicable functionalities reported by UE. Otherwise the step of reporting applicable functionalities seems useless. |
| Mediatek | No | I tend to agree with other companies that we don’t need to couple configurability and applicability together. It also relates to whether proactive or reactive reporting for UE-side applicable functionality is assumed. |
| Kyocera | Others | For the question in Q5, our answer is “Yes.” However, we are wondering if we need to discuss it further, since configuration and applicability are different topics. |
| Huawei, HiSilicon | See comments | Related to Q2.  **We suggest to clarify the content and the purpose of the configuration for the terminology "Configured functionalities" before discussing Q5.** |
| Intel | Depends | In our understanding, this depends on when and how the configuration is provided to the UE. Furthermore, the definition of configured functionalities need to be clarified first according to Q4. |
| LGE | See comments | In the case of a reactive report, the configured functionality might not be immediately applicable (like CHO target cell configuration), and in this sense, I agree with rapp. |
| Sharp | No | Configured functionality are applicable and hence applied but not all the applicable functionalities may be configured.   * **Configured Functionality:** This refers to functionalities set up by the network (gNB) through RRC configuration. It's a network-driven action. * **Applicable Functionality:** This depends on whether the functionality is actually usable by the UE at a given time. It considers the UE's capabilities (current state, resources, model availability) and any additional network conditions. This is a UE-driven decision.   The key point is that configuration doesn't guarantee applicability. The UE might not be able to use a configured functionality due to its own limitations or external factors. These are two independent concepts based on different timeframes:   * Configuration happens during RRC procedures. * Applicability is determined based on dynamic changes in the UE and network environment. |
| DOCOMO |  | See our comment to Q2. It is unclear what timing of configuration is defined by “configured functionality”. |
| Qualcomm | Yes | We do not see an issue with the definition. There are only two possibility   * Configured functionalities can be activated immediately (configuration and activation happens together). * Configured functionalities cannot be activated immediately (configuration and activation happens separately).   We do not see a third possibility. So, we do not see any issue with the definition provided by rapp. |

## Activated functionalities

|  |
| --- |
| *Activated functionalities:* this refers to functionalities that the UE starts predicting beam results via model inference. |

**Q6: Do you agree that activated functionalities refer to the functionalities that the UE uses beam prediction/positioning via model inference?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | partially Yes | We think “**UE uses beam prediction/positioning”** is not necessary, especially considering CSI compression/prediction is still on the table of Rel-19. Since it is a high level definition, we prefer to make it more general, e.g.  **activated functionalities refer to the functionalities that the UE uses ~~beam prediction/positioning via~~ output of model inference** |
| vivo | Yes | Can be refined as  *Activated functionalities:* **this refers to functionalities that have been enabled for ~~the UE starts predicting beam results via~~ model inference.** |
| OPPO | partially Yes | The similar view as Apple, let’s make it more generic for now.  **activated functionalities refer to the functionalities that the UE is using ~~beam prediction/positioning via~~ the output from model inference** |
| Xiaomi | Yes |  |
| ZTE | Yes | We can make it more general like apple, vivo, oppo’s suggestion. |
| Ericsson | Yes, see comment | Agree with the intention and to make it a general definition (i.e., extend it to positioning). Here a proposed rewording:  *“Activated functionalities:* refers to AI/ML functionalities already activated and performing inference.” |
| NEC | Yes | It would be good to have a generic definition to cover all cases. No strong view on the detailed wording. |
| Lenovo | Yes | The wording can be modified as companies suggested above. |
| CATT | partially Yes | Agree with the comments above to make it a general definition, and Ericsson’s wording looks good to us. |
| Mediatek | Yes with comment | Agree with the intention. Activated functionalities refer to the AI/ML functionalities that are currently in use for inference and have been activated by the network. |
| Kyocera | partially Yes | We prefer to change it to be more generic, i.e., Activated functionalities refer to the functionalities where the UE has explicitly enabled model inference. |
| Huawei, HiSilicon | Yes with comments | OK with vivo's suggestion. |
| Intel | Yes | Changes from Ericsson looks good to us. |
| LGE | Yes with comments | OK with Vivo’s change and Ericsson’s change |
| Sharp | Yes, with comments | Agree with the companies, it can be more generic |
| DOCOMO | Yes with comment | It is ok to make it more general.  However, it is unclear whether this definition includes for functionality activation to monitor performance or not. |
| Qualcomm | Yes with comment | Prefer to have  “Activated functionalities refer to the AI/ML functionalities that have been activated by the network” |

## Available functionalities

During RAN2 discussion, there is a proposal on availability/available functionalities [2].

|  |
| --- |
| *Proposal 2 Introduce signalling for the UE to inform the gNB whether the AI/ML functionality is available for operation (e.g., whether there are trained models available within it). FFS whether the “availability indication” can be reported as part of the applicability-reporting information, or as a separate signalling.* |

**Q7: Do you agree that available functionalities should be separately considered from applicable functionalities?**

If yes, please comment what separate characteristics should be added/expected compared to applicable functionalities.

If no, please suggest how to consider available functionalities in RAN2 discussion (e.g. consider same as applicable functionalities or postpone to the next meeting)

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | No (same as applicable functionalities) | Please note that above proposal was agreed as different wording after online discussion:  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.  So, our understanding is that above agreement (esp. the highlighted part) has implied “applicable functionality” is same as “available functionality”.  Meanwhile, we also understand that the highlighted part implies the definition of “applicable functionalities suggested by moderator:  **applicable functionalities refer to functionalities that the UE is ready to apply for model inference and they can be considered as candidates for functionality activation/deactivation** |
| vivo | No | The available functionality is the same as the applicable functionality.  The intention to have separate definitions is not clear. |
| OPPO | Yes | Based on our picture raised in Q2, we understand available functionalities are the super set of applicable functionalities, not all available functionalities are applicable based on NW additional condition and UE inside additional condition, so we can have the following simple definition for available functionalities:   * **available functionalities refer to functionalities that the UE has the corresponding model(s).** |
| Xiaomi | No | The definition and usage of available functionality are not clear. Maybe we can further discuss in next meeting based on contributions if needed. |
| ZTE | Yes | We agree to have this definition for the future discussion, in our understanding, for one functionality reported as supported at UE side via UE capability, the main concern from NW is whether there is any available models for such functionality, whether the functionality is applicable to be activated at the time being. In this sense, only available functionality and applicable functionality does the matter for NW. So we support to have a clarification of the definition in terms of available functionality, the following is a suggestion on top of OPPO’s version:   * **available functionalities refer to functionalities that the UE has the model(s) to perform the inference.** |
| Ericsson | Yes | In principle, availability is not the same as applicability. This can also be extracted from RAN2#126’s agreement:  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.  As per the agreement, available does not mean applicable, while applicable means available.  In our understanding it is important to distinguish this, especially if the UE reports that a functionality is not applicable. As ZTE pointed out, if the functionality is not applicable, but the model is available, then the gNB can provide an RRC configuration such that the model becomes applicable (i.e., inference configuration that fits the trained data set). But if the model is completely unavailable, then the gNB does not have any possibility to provide a suitable inference configuration.  So at least from a stage-2 level, there is a clear distinction between applicability and availability, and it is worthwhile clarifying that. |
| NEC | Yes | We understand that “available functionality” refers to supported functionalities matching the UE side additional conditions whereas “applicable functionality” refers to available functionality matching the NW side additional conditions. Therefore, we shame the same view as OPPO that applicable functionality is a subset of available functionality.  Moreover, we don’t think the UE needs to report available functionalities to the NW, only reporting applicable functionality is sufficient. |
| Lenovo | Maybe No | For an applicable functionality it must has the model available as agreed during the meeting.   * For a functionality to be applicable at least there should at least one model available within it.   Besides, in our understanding, it would be reasonable to assume that for a supported functionality, at least one model for this functionality is available at UE already.  Thus, maybe we don’t need a concept of “available functionality” at the end, as it is covered in the “supported functionality” and “applicable functionality”. |
| CATT | No | In our opinion, only the applicable functionalities fulfilling the UE and NW additional conditions need to be reported to the NW for configuration/activation.  So in the stage 3 spec, it’s sufficient that only the applicable functionalities need to be clarified:  **applicable functionalities refer to functionalities that the UE is ready to apply for model inference and they can be considered as candidates for functionality activation/deactivation**  WE think “is ready to” here also covers the meaning that the model is available to be used by UE. |
| Mediatek | No | We agree with OPPO that available functionalities are a superset of applicable functionalities, meaning not all available functionalities are applicable. However, the question remains as to why or whether the UE needs to report the available functionalities to the network. We believe that availability information should be retained by the UE and does not need to be communicated to the network.  Given the agreement that the UE determines the applicability of a functionality, when the UE sends an applicability indication to the network, it inherently implies that the functionality already has an available model. |
| Kyocera | No | In our understanding, available functionality is quite similar to applicable functionality. The difference between them is simply that “at least one model should be available within it.” We understand this is necessary for signalling efficiency. Therefore, in order to simplify the types of functionalities, we think that available functionality is unnecessary. |
| Huawei, HiSilicon | Yes with comments | Firstly, we think RAN2 can discuss this terminology "available functionlities", and ZTE's definition can be used as the starting point. For this definition, it means the model(s) may be or may not be applicable in the present circumstances.  Secondly, we need to check whether the UE should tell such information to the NW side or not, and the following questions can be clarified:  (1) what information is to be sent from UE to NW  (2) what is the use of the Information at NW side, and what is the next step for NW  For our above considerations, we think RAN1 involvement may be needed. For example, if UE has models but they are not applicable due to NW-side additional conditions, whether the NW needs to know this situation at UE side and what NW can do with it.  Also, for the following comment from Ericsson, we may also need to check it with RAN1.  if the functionality is not applicable, but the model is available, then the gNB can provide an RRC configuration such that the model becomes applicable (i.e., inference configuration that fits the trained data set). |
| Moderator | See comments | Approach 1:   * Applicable functionalities are from the NW point of view that the functionalities are ready for model inference (i.e. candidates for activated functionalities). * Available functionalities are from UE point of view that the functionalities are ready for model inference (i.e. have available models).   Approach 2:   * Use one terminology: Applicable functionalities are functionalities ready for model inference from both the NW and the UE point of views. [gNB may have subset of applicable functionalities that UE have. ]   I wonder if we can start with approach 1 and RAN2 can discuss later whether both terminologies should be specified or one/none need to be specified after discussion on the detailed procedure. |
| Intel | Yes | Agree there’s a difference between available functionalities and applicable functionalities. Available functionalities may or may not be applicable, where the available functionalities may become applicable after network provides the proper configurations according to associated to training.  We are ok with the definition from Oppo, which seems generic. Considering the functionalities may further require configuration for model training data collection, we think it would be better not to limit it to model inference only.  We are ok to go with approach 1 with below definition:  **applicable functionalities refer to functionalities that the UE is ready to apply for model inference**  **available functionalities refer to functionalities that the UE has the corresponding model(s).** |
| LGE | No with comments | I agree with Apple's comment(support Approach 2).  However, since inference can be determined by the presence of a model, the existence of relevant configurations, and whether UE/NW sided conditions are met, it may be necessary for the NW to understand what is not applicable for the inference. We can rediscuss it from a procedural perspective, not in terms of terminology. |
| Sharp | No | Applicable functionality and available functionality can be merged and considered in a simplified way. The intention to separately consider applicable and available functionality is not clear.  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.  If there is no model available, the functionality may still be **‘made’** applicable if the model can be acquired through a model transfer/delivery mechanism.  As mentioned in Q1 response. The terms Functionality applicability and model availability should not be mixed and needs to be considered separately. |
| DOCOMO | See comment | We support moderator’s intention, i.e., RAN2 can discuss later.  We understand the difference between available functionalities and applicable functionalities, but we are not sure the necessity and benefit to separate these definitions. |
| Qualcomm | No | UE determines applicable functionalities based on model availability at the UE, network side additional conditions, and UE side additional conditions. We agree with Mediatek that UE does not need to indicate model availability information to the network, “availability information should be retained by the UE and does not need to be communicated to the network.”  The UE reports applicable functionalities beyond the configured functionalities, as this will allow the network to reconfigure the UE without further message exchanges. |

# Phase 2

# Conclusion

# Reference

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

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