3GPP TSG RAN WG1 #113 **R1-2304476**

Incheon, Korea, May 22nd – May 26th, 2023

Source: **vivo**

Title: **Other aspects on AI/ML for positioning accuracy enhancement**

Agenda Item: **9.2.4.2**

Document for: **Discussion and Decision**

1. Introduction

Many positioning methods have been specified in Rel-16 and Rel-17 NR positioning, to obtain position estimation with target horizontal positioning accuracies of <0.2 m (90%) for IIoT use cases and <1 m (90%) for commercial use cases. However, the performance of these positioning methods highly relies on the existence of multiple LOS (line-of-sight) paths between the target terminal and multiple TRPs (Transmission-Reception Points). In the scenarios with extremely low LOS probability, positioning accuracy would decrease dramatically, which may be not able to satisfy the high-accuracy positioning requirements coming from new applications and industry verticals.

AI/ML technology has powerful abilities in feature extraction, environment awareness, complex problem modeling and processing. In recent years, applying AI/ML into air-interface has attracted great attentions from academics to industries, and a lot of meaningful exploration has been made to verify the performance gain compared to conventional non-AL/ML schemes. Related research has also verified that AI/ML technology has the potential to significantly improve the performance of wireless communications.

Under this background, a new SI on *Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface* has been agreed at RAN #94e [1], including three use cases to assess the applications of AI/ML in air-interface. Among them, AI/ML based positioning accuracy enhancement is included, with the target to improve the positioning accuracy for different scenarios, especially for some challenging scenarios with heavy NLOS (non-line-of-sight) conditions.

The objective of the new SI for RAN1 AI/ML based positioning includes the following:

> Study the 3GPP framework for AI/ML for air-interface corresponding to each target use case regarding aspects such as performance, complexity, and potential specification impact.
> Use cases to focus on:
> 1. Initial set of use cases includes:
> b. Positioning accuracy enhancements for different scenarios including, e.g., those with heavy NLOS conditions [RAN1]

All agreements reached in previous meetings can refer to Appendix A. In this contribution, we present our views on sub use cases and potential specification impact for AI/ML based positioning accuracy enhancement.

2. Potential specification impact

In this section, we analyze the potential specification impact of AI/ML based positioning accuracy enhancement from various perspectives, including model training, model indication/configuration, model monitoring and update, model inference and required UE capability report.

2.1. Common framework

According to the common functional framework for RAN intelligence illustrated in Figure 1 [3], we analyze the general specification impact from the perspective of lifecycle management for AI/ML based positioning accuracy enhancement.

- Data Collection is a function that provides a training dataset to model training and input data to model inference functions. An example of a training data format is {CIR, location}, and accordingly, the input data to model inference functions is CIR. Especially, a data pre-processing procedure is necessary after data collection, such as CIR truncation and translation. Data normalization has the potential to decrease the overhead for data collection and improve the model performance.
- Model Training is a function that performs the AI/ML model training, validation, and testing which may generate model performance metrics as part of the model testing procedure. AI/ML model training can be

further divided into two phases, i.e., model pre-training based on collected data and model update based on collected field data if the current model does not work well.

- Model Inference is a function that provides AI/ML model inference output (e.g. predictions or decisions). In different sub use cases, the output of model inference could be different, such as UE's location and an intermediate feature.
- Actor is a function that receives the output from the model inference function and triggers or performs corresponding actions. For AI/ML assisted positioning, the output from the Model Inference function is the intermediate feature, and then Actor performs position estimation according to the intermediate feature.

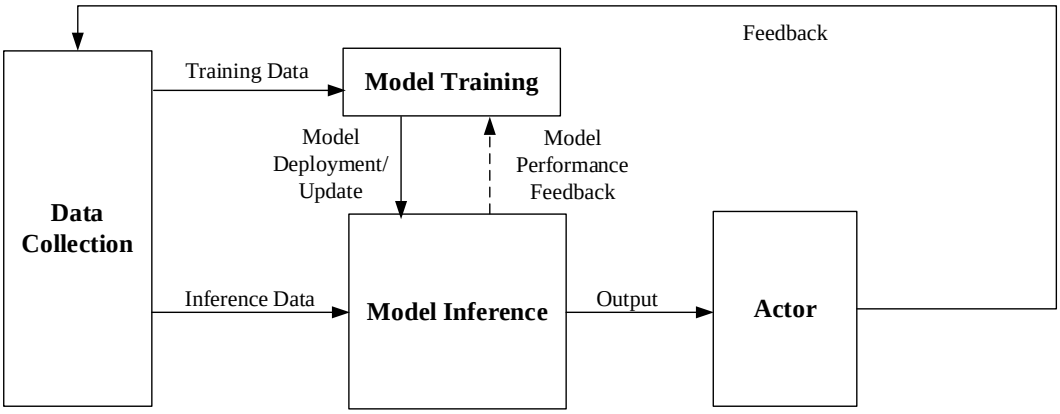

Figure 1 A common functional framework for RAN intelligence [3]

2.2. A general discussion for various positioning cases

At the RAN1#110-bis meeting, it was agreed that:

Agreement

- Study and provide inputs on benefit(s) and potential specification impact at least for the following cases of AI/ML based positioning accuracy enhancement
  - Case 1: UE-based positioning with UE-side model, direct AI/ML or AI/ML assisted positioning
  - Case 2a: UE-assisted/LMF-based positioning with UE-side model, AI/ML assisted positioning
  - Case 2b: UE-assisted/LMF-based positioning with LMF-side model, direct AI/ML positioning
  - Case 3a: NG-RAN node assisted positioning with gNB-side model, AI/ML assisted positioning
  - Case 3b: NG-RAN node assisted positioning with LMF-side model, direct AI/ML positioning

We further make a clarification for the sake of subsequent discussions, which is as listed in Table:

Table 1 A detailed list of various positioning cases

Cases	Model deployment	Positioning methods	Measured by which entity	Model output	Position calculation
Case 1	UE-side	Direct AI/ML positioning	UE (Downlink positioning)	Position	UE-side
		AI/ML assisted positioning		Intermediate feature	
Case 2a	UE-side	AI/ML assisted positioning	UE (Downlink positioning)	Intermediate feature	LMF
Case 2b	LMF-side	Direct AI/ML positioning	UE (Downlink positioning)	Position	LMF
Case 3a	gNB-side	AI/ML assisted positioning	gNB (uplink positioning)	Intermediate feature	LMF

Case 3b	LMF-side	Direct AI/ML positioning	gNB (uplink positioning)	Position	LMF

2.3. Data collection for model training

The high-accuracy positioning relies on a well-trained AI/ML model deployed at UE side or network side. In particular, training data is the most important part for model training, and high-quality data collection is a prerequisite to enable such model training. To guarantee the quality of data collection, some assistance information interaction between UE side and network side may be necessary. In this section, we mainly discuss the potential specification impacts from the perspective of data collection.

AI/ML is data-driven, and data collection is a necessary but challenging process for AI/ML based positioning. As shown in our companion contribution [2], time domain channel CIR as the input of AI/ML model can obtain the best positioning accuracy compared to other inputs, such as power, delay and angle of the first path. Therefore, we propose

Proposal 1: Support time domain CIR as one model input for training of AI/ML model for positioning.

At the RAN1#111 meeting, it was agreed that:

Agreement

Regarding data collection for AI/ML model training for AI/ML based positioning,

- The following options of entity and mechanisms to generate ground truth label are identified for further study
  - For direct AI/ML positioning, ground truth label is UE location
    - PRU with known location
    - UE generates location based on non-NR and/or NR RAT-dependent positioning methods
    - LMF generates UE location based on positioning methods
    - LMF with known PRU location
    - Note: user data privacy needs to be preserved
  - For AI/ML assisted positioning, ground truth label is one or more of the intermediate parameter(s) corresponding to AI/ML model output
    - PRU generates label directly or calculates based on measurement/location
    - UE generates label directly or calculates based on measurement/location
    - Network entity generates label directly or calculates based on measurement/location
- The following options of entity to generate other training data at least measurement corresponding to model input are identified for further study
  - For UE-based with UE-side model (Case 1) and UE-assisted positioning with UE-side (Case 2a) or LMF-side model (Case 2b)
    - PRU
    - UE
  - For NG-RAN node assisted positioning with Network-side model (Case 3a and Case 3b)
    - TRP
  - Note: other options of entity to generate other training data are not precluded
- Note: Existing PRU definition is in 38.305

Agreement

Regarding data collection for AI/ML model training for AI/ML based positioning, study benefits, feasibility and potential specification impact (including necessity) for the following aspects

- Request/report of training data
  - Ground truth label
  - Measurement corresponding to model input
  - Associated information of ground truth label and/or measurement corresponding to model input
- Assistance signaling and procedure to facilitate generating training data
  - Reference signal (e.g., PRS/SRS) configuration(s) and configuration identifier
  - Assistance information, e.g., between LMF and UE/PRU, for label calculation/generation, and label validity/quality condition, etc.
  - Note1: whether such assistance signaling and procedure can be applied to other aspect(s) of AI/ML model LCM can also be discussed

- Note2: Study may consider different entity to generate training data as well as different types of training data when applicable
- Note3: study considers both of the following cases when applicable
    - when the training entity is the same entity to generate training data
    - when the training entity is not the same entity to generate training data

At the RAN1#112 meeting, it was agreed that:

Agreement

Regarding training data generation for AI/ML based positioning,

- The following options of entity and mechanisms to generate ground truth label are identified
    - At least PRU is identified to generate ground truth label for UE-based positioning with UE-side model (Case 1) and UE-assisted positioning with UE-side model (Case 2a)
    - At least LMF with known PRU location is identified to generate ground truth label for UE-assisted/LMF-based positioning with LMF-side model (Case 2b) and NG-RAN node assisted positioning with LMF-side model (Case 3b)
    - At least network entity with known PRU location is identified to generate ground truth label for NG-RAN node assisted positioning with gNB-side model (Case 3a)
    - FFS whether and if so, applicable conditions and potential specification impact for the following options to generate ground truth label
        - UE generates ground truth label based on non-NR and/or NR RAT-dependent positioning methods
        - Network entity generates ground truth label based on positioning methods
- The following options of entity to generate other training data (at least measurement corresponding to model input) are identified
    - For UE-based with UE-side model (Case 1) and UE-assisted positioning with UE-side (Case 2a) or LMF-side model (Case 2b)
        - PRU
        - UE
    - For NG-RAN node assisted positioning with Network-side model (Case 3a and Case 3b)
        - TRP
- Note: transfer of training data from the entity generating training data to a different entity is not precluded and associated potential specification impact is for further study

Agreement

Regarding training data collection for AI/ML based positioning, study benefit(s) and potential specification impact (including necessity) at least for the following aspects

- Associated information of training data
    - Quality indicator at least for ground truth label (if needed)
    - Other information associated with training data is not precluded. E.g., information related training dataset/samples, information related to scenario, resource configuration & mapping, timing for training data, information on implementation imperfections, etc.
- Assistance signaling and procedure to facilitate generating/collecting training data
    - Potential determination of the UE/PRU/TRP which can provide the training data
    - Configuration of reference signal (for measurement and/or label)
    - Signaling other than above 2 for data collection
        - E.g., requested quality of training data

At RAN1#112bis-e, the following working assumption was reached.

Working Assumption

Regarding data collection at least for model training for AI/ML based positioning, at least the following information of data with potential specification impact are identified.

- Ground truth label
    - At least for model training
    - Report from the label data generation entity
- Measurement (corresponding to model input)
    - At least for model training
    - Report from the measurement data generation entity
- Quality indicator
    - For and/or associated with ground truth label and/or measurement at least for model training

    - Report from the label and/or the measurement data generation entity and/or as request from a different (e.g., data collection, etc.) entity
- RS configuration(s)
    - At least for deriving measurement
    - Request from data generation entity (UE/PRU/TRP) to LMF and/or as LMF assistance signaling to UE/PRU/TRP
    - Note1: there may not be any enhancements on top of existing RS configuration(s) or any new RS configuration(s) for positioning measurement
- Time stamp
    - At least for and/or associated with training data for model training
        - Separate time stamp for measurement and ground truth label, when measurement and ground truth label are generated by different entities
    - Report from data generation entity together with training data and/or as LMF assistance signaling
    - Note2: there may not be any enhancements on top of time stamp in existing positioning measurement report or any new time stamp report for positioning measurement
- FFS other necessary information (e.g., scenario identifier. LOS/NLOS condition, timing error, etc.) for data collection
- Note3: whether the above information can be applied to other aspects of AI/ML LCM (e.g., updating, monitoring, etc.) can also be discussed
- Note4: transfer of data from the entity generating data to a different entity is not precluded from RAN1 perspective

In general, for downlink positioning, the input for an AI/ML model training is CIR estimated from downlink PRS measurement plus UE ground truth location data. Such training data including training input and labels, could be collected by UE side, including PRUs and/or regular UEs. Among them, PRUs deployed in advance are more convenient to collect data with ground truth labels, and regular UEs may be also able to collect some data with noisy labels or without labels. For uplink positioning, the input for an AI/ML model training is CIR estimated from uplink SRS measurement plus UE ground truth location data. Such training data could be collected by the collaboration of network side and UE side. For example, ground truth label could be collected by UE side, and training input, such as CIR, should be collected by network side.

Some dedicated reference signal configuration may be also required to support data collection. For AI/ML based positioning, some non-ideal factors can severaly impair the positioning performance, including synchronization error, timing error, CIR estimation error, and labeling error. To mitigate their impacts, the technologies from the AI/ML perspective, such as mixtraining and fine-tuning, has been proposed. Besides, RS configuration enhancement may also a feasible solution to mitigate or eliminate the impact of these factors, such as configuring RSs to estimate synchronization error and interference and further compensate their impacts .

In this contribution, only centralized learning (model training at a single side) is assumed without considering distributed learning. When the entity conducting model training and data collection is not the same, collected data should be delivered from the data-collection entity to the model-training entity.

Proposal 2: When the entity conducting model training and data collection is not the same, collected data should be delivered from the data-collection entity to the model-training entity.

In the following sub-sections, we further analyze the procedures and specification impacts for different cases as agreed in RAN1#110-bis meeting.

2.3.1. Case 1 and Case 2

According to Table 1, Case 1 and Case 2 belong to downlink positioning, meaning that training input should be measured and collected at UE side. Moreover, primary training label including location coordinate or other intermediate feature should be collected by UE side. Optionally, indirect training label can be also obtained by NW side after the primary training label is delivered to NW side and then pre-processed. For example, location coordinate is collected by UE side, and related TOA labels can be calculated at NW side by combining the location coordinates of UE and gNBs.
Here, we will discuss the specification impacts for Case 1 and Case 2 in detail.

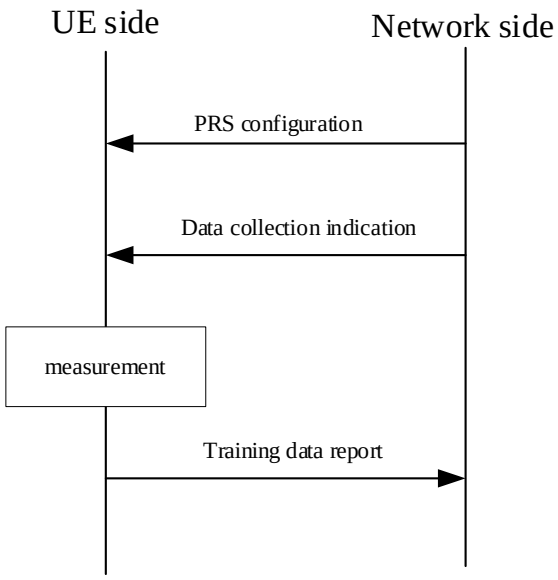

Figure 2 The procedure of data collection for data measurement at network side

A possible procedure is shown in Figure 2, when model training is performed at NW side, and the details are clarified as follows:

- Network side sends PRS configuration signaling to the target UEs when a dedicated PRS configuration is required to support data collection.
- Network side sends a data collection indication to the target UEs, which may consist of training data size and type, data preprocessing, and so on.
- The target UEs perform data measurement, and then report collected data (including training input and training labels) to network side, where dedicated resources for data report should be allocated in advance.

When model training is performed at UE side, training data report from UE side to NW side is no longer necessary.

2.3.2. Case 3

According to Table 1, Case 3 belongs to uplink positioning, meaning that training input should be measured and collected by NW side. Specifically, training input should be measured by gNB side and collected by LMF side. Moreover, primary training label including location coordinate or other intermediate feature should be collected by UE side. Optionally, indirect training label can be also obtained by NW side after the primary training label is delivered to NW side and then pre-processed. For example, location coordinate is collected by UE side, and related TOA labels can be calculated at NW side by combining the location coordinates of UE and gNBs.

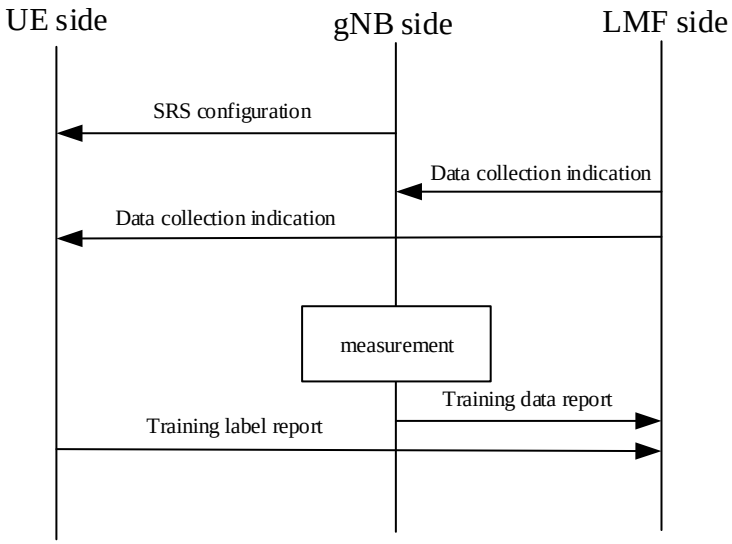

Figure 3 The procedure of data collection for data measurement at network side

As shown in Figure 3, when model training is performed at NW side, the detailed specification impacts are listed as follows:

- LMF side sends data collection indication to the target UEs and gNB to perform UL measurement.
- gNB side provides measurement results for training input, such as CIR.
- The target UEs report related training labels to network side, where dedicated resources for training labels report should be allocated in advance.

When model training is conducted at UE side, training input collected by NW side should be delivered to UE side, and training label report is no longer necessary. However, considering privacy protection and data transfer overhead, this framework of UE training for Case 3 should be avoided.

2.3.3. Assistance information

To support high-quality data collection, we think related assistance information at least consists of RS configuration and data collection indication. Two examples are presented as follows:

- **RS configuration**: As presented in our companion contribution [2], it is observed that some imperfect factors can severely impair positioning accuracy. For example, the interference from other TRPs can severely deteriorate channel estimation quality, and further impair positioning accuracy of AI/ML. In this sense, some dedicated RS resources should be configured to eliminate the interference across TRPs so as to ensure the data quality.
- **Data collection indication**: As presented in our companion contribution [2], training label error can also impair the positioning accuracy of AI/ML model. To improve labeling quality, it is necessary to indicate the criteria for data label filtering to discard the training samples with low-confidence labels. Indicating UE to report label quality indicator is also optional. Moreover, other indications such as measurement type and format, should be included. Specifically, measurement type indicates what kind of information should be measured for positioning, such as CIR or PDP. Measurement format indicates the number of time-domain samples of CIR, the arrangement of CIRs of different TRPs, and other possible pre-processing. In particular, the measurement type and format should be dependent on model input.

Proposal 3: For ground truth label collection, to improve the quality of labels, indicate UE the criteria or requirement for data labeling or indicate UE to report label quality indicator.

Proposal 4: For measurement collection, at least measurement type and format should be indicated, depending on the model input.

2.3.4. Entity for data collection

As presented in our companion contribution [2], training an AI/ML model with limited labeled data is feasible by utilizing the proposed semi-supervised learning framework. Specifically, in addition to labeled data, large amounts of unlabeled data which is easy to collect in practice, can also be used to enhance the performance of model training. In this way, the cost of data collection can be mitigated significantly. Both PRUs and regular UEs can participate in the data collection process, which may greatly reduce the time of data collection and accelerate model deployment in practice. Thus, it is meaningful to further study the impact of data collection for semi-supervised learning.

In previous agreement, only PRU is used to generate ground truth label. During RAN1#112bis-e, there's a discussion on whether to use UE to generate ground truth label at least for case 1 and 2a. Some companies do not support to use UE to generate ground truth label for the concern of performance loss compared to the ideal ground truth label. However, we don't think such argement is relevant as the motivation to use UE to generate ground truth label is to solve the issue where there may not enough PRU for data collection. As we presented in our companion contribution [2], evalution results show that even for noisy label estimated by UE with legacy positioning methods, performance accuracy benefits can still be obtained when those noisy labels with associated label quality indicator were reported and used for model training.

Therefore, both PRUs and regular UEs shoul be used to perform data collection. Considering the different terminal capabilities, data collection requirements may be different for different types of terminals. For example, PRU can be used to collect training input and related label (e.g., location) at the same time, while regular UE can be used to collect noisy label and/or training input (e.g., CIR measurement) even if related training label (e.g., UE location) is not available. In such case, some additional information such as data collection for supervised learning or semi-supervised learning, with label or without label, may be included into the data collection indication. Moreover, these unlabeled data collected from regular UE can also be used to perform model monitoring, such as model monitoring based on data distribution of CIR.

Proposal 5: In addition to PRU, UE is also used to perform data collection for both label and measurement.

Proposal 6: Further study the specification impact of data collection for semi-supervised learning.

2.3.5. Other potential issues

Some potential issues are listed as follows:

- When the training entity is not the same entity to obtain training label and/or other training input, the corresponding mechanism should be studied to ensure the matching between training label and training input. For example, additional time stamp label needs to be marked during training input and label measurement.
- Considering limited battery capacity and available training data of mobile terminals, online model training with a large-scale dataset should be avoided at UE side since model training usually requires large amounts of training data and computational resources. It is worth noted that AI/ML model updating with a small dataset at the UE side should be supported and should not be excluded as type of UE online training.
- Considering different terminal capabilities, a unified data quality requirement may be necessary to ensure data-collection quality.

Proposal 7: Real-time on-device model training with a large-scale dataset should be avoided at UE side.

Proposal 8: Other potential issues on data quality and terminal capability should be considered for data collection.

2.4. Model indication/configuration

To support flexible model deployment, some necessary assistance information related to model indication is required. In general, model indication consists of signaling for model transfer, model activation/deactivation, and model selection.

2.4.1. Model transfer

At the RAN1#110 meeting, it was agreed that:

Agreement

Study aspects in terms of potential benefit(s) and requirement(s)/specification impact(s) of AI/ML model training and inference in AI/ML for positioning accuracy enhancement considering at least

- UE-side or Network-side training
- UE-side or Network-side inference
  - Note: model inference at both UE and network side is not precluded where proponent(s) are encouraged to clarify their AI/ML approaches

Following the above agreement, there are five possible cases for model training and inference:

- Case 1: Model training at network side, model inference at UE side.
- Case 2: Model training at UE side, model inference at network side.
- Case 3: Model training at some UEs, model inference at other UEs (requires model transfer, e.g., different UE manufacturers).
- Case 4: Model training and inference at network side.
- Case 5: Model training and inference at UE side (requires no model transfer).

As shown in Table 2, we further analyze the advantages and disadvantages of these cases in detail from the perspectives of overhead and performance.

Table 2 Analysis of advantages and disadvantages for different cases of model training/inference

Case 1: Model training at network side, model inference at UE side.	
Need model transfer?	Yes
Advantages	Disadvantages
• Less concern on training complexity due to the availability of enough training hardware resource at NW. • Model life cycle management (e.g., monitoring and update) can be better supported by NW since the full knowledge of the model is available at NW.	• The overhead of model delivery cannot be neglected. • A unified AI/ML model representation format is required to align the understanding of the model across network and UEs. • The overhead of training data reporting cannot be neglected.

Case 2: Model training at UE side, model inference at network side	
Need model transfer?	Yes
Advantages	Disadvantages
• The computational resource of UE side can be leveraged to train an AI model jointly through distributed learning frameworks, such as federated learning. • The overhead of training data reporting can be avoided.	• The AI/ML model trained by UE side may not be always reliable or suitable, and applying distributed learning into communication is premature and still has a long way to go. • The overhead of model delivery cannot be neglected. • A unified AI/ML model representation format is required to align the understanding of the model across network and UEs.
Case 3: Model training at some UEs, model inference at other UEs (requires model transfer)	
Need model transfer?	Yes
Advantages	Disadvantages
• The computational resource of some UEs can be leveraged to train an AI model jointly through distributed learning frameworks, such as federated learning. • A wide range of other UEs (e.g., those not capable of training due to resource constraints and/or from different UE manufactures) can still utilize that trained model for system performance benefit. • The overhead of training data reporting can be avoided.	• The overhead of model delivery can be not neglected. • A unified AI/ML model representation format is required to align the understanding of the model across network and UEs.
Case 4: Model training and inference at network side	
Need model transfer?	No
Advantages	Disadvantages
• The overhead of model delivery is negligible • A unified AI/ML model representation format across network and UEs is not required. • Less concern on training complexity due to the availability of enough training hardware resource at NW. • Model life cycle management (e.g., monitoring and update) can be better supported by NW since the full knowledge of the model is available at NW.	• The overhead of training data collection cannot be neglected. • The overhead of training data reporting cannot be neglected if AI/ML model relies on UE measurement/report as the input
Case 5: Model training and inference at UE side (requires no model transfer)	
Need model transfer?	No
Advantages	Disadvantages
• The overhead of model delivery is negligible • A unified AI/ML model representation format across platforms is not required.	• Only limited UEs (e.g., those with training capability and/or from the same UE manufacture as those training capable UEs) can utilize the AI/ML model • The AI/ML model trained by UE side may not be always reliable and/or suitable for

	deployment scenario due to the limited training data and hardware resource. • It is hard for NW to help UEs with model monitoring and update due to the lack of model meta-information at NW.

As we discussed above, online model training with a large scale dataset should be avoided at UE side. Then at least for downlink positioning, pre-trained AI/ML model should be indicated to UE side such that a UE can conduct inference to get the UE location.

When AI/ML model is deployed at network side, there is no specification impart on air-interface for model transfer. When AI/ML model is deployed at UE side, network side should transfer the model information to the target UE. The model information may consist of the input and output of model, the architecture of the model, the weight of the model, the configuration of the model, the state of the optimizer, and so on. Moreover, UE side should report the device capability to network side in advance, so that network side can verify whether the target UE supports AI/ML based positioning or what kind of model should be configured for the target UEs. The detailed specification impacts are listed as follows.

- UE side reports the UE capability to network side. This process may also be triggered by network side, such as sending a device capability request to UE side.
- Network side sends the model information to the target UE if the target UE can support AI/ML based positioning.

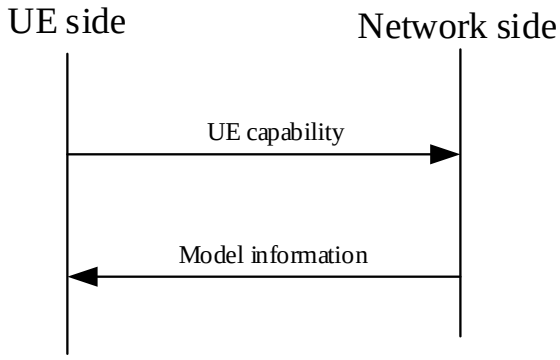

Figure 4 Information exchange procedure for model configuration

For model transfer, there are three types in general:

- Type 1: Transfer model parameter only.
- Type 2: Transfer model structure only.
- Type 3: Transfer both model parameter and model structure.

At the current stage, at least Type 1 should be primarily considered. To estimate the signaling overhead of Type 1, we initially evaluate the amount of parameters that need to be transmitted for acceptable positioning performance. As shown in Table 1, it is observed that high-accuracy positioning can also be achieved by these AI/ML models with extremely small parameter scale. For example, when only 6k model parameters are transferred over air interface, the positioning accuracy of 2m@90% is still maintained. Moreover, the bits transmitted over air interface can be further reduced by model quantization technologies. In theory, float16 quantization can reduce 50% bits overhead required for model parameter transfer as compared with float32 quantization, and int8 quantization can reduce 75% bits. We can observe that there is no performance degradation for float16 quantization. Int8 quantization has a slight performance degradation, but the positioning accuracy is still acceptable. It is worth noting that some methods can further reduce the number of bits for model quantization on the premise of ensuring the performance of AI/ML model, and thus the overhead of model transfer can be significantly reduced.

Table 1 Comparison of different numbers of model parameters for different model quantization schemes

Number of model parameters	Positioning accuracy (m) @90% CDF of various model quantization schemes		
	Float 32	Float 16	Int 8

34k	0.95	0.95	1.59
21k	1.14	1.14	2.21
9k	1.40	1.40	2.20
6k	2.12	2.12	3.47

Observation 1: Model transfer over air interface can be achieved with extremely low signaling overhead by combining small-parameter model design and advanced model quantization technologies.

Proposal 9: Further study the overhead of model transfer, and support model transfer over air interface for AI/ML based positioning.

At the RAN1#110-bis meeting, it was agreed that:

Agreement

Regarding AI/ML model indication[/configuration], to study and provide inputs on potential specification impact at least for the following aspects on conditions/criteria of AI/ML model for AI/ML based positioning accuracy enhancement

- Validity conditions, e.g., applicable area/[zone/]scenario/environment and time interval, etc.
- Model capability, e.g., positioning accuracy quality and model inference latency
- Conditions and requirements, e.g., required assistance signalling and/or reference signals configurations, dataset information
- Note: other aspects are not precluded

In particular, model information should contain meta-information. The function of meta-information is to indicate the model capability and the physical and network environment or condition under which the model is suitable for operation. Specifically, such meta-information may consist of:

- **Validity condition:** indicating the applicable area/zone/scenario/environment of the associated AI/ML model. For example, the target UE should switch to a new model when out of the current model's applicable area.
- **Model capability**: including model accuracy achieved on training verification and/or test dataset, model inference latency. For example, the target UE should switch to a new model when the current model cannot meet the positioning accuracy or inference latency requirements.
- **Conditions and requirements**: indicating model-related RS configuration, SINR range, synchronization error range and so on. For example, the target UE should switch to a new model when the measured SINR is out of the SINR range of training dataset.

Proposal 10: For the case where model is developed at network side and deployed at UE side, network side should transfer the model information to the target UE.

Proposal 11: Model information should contain meta-information indicating model capability and the physical and network environment or condition under which the model is suitable for operation.

2.4.2. Model activation/deactivation

An AI/ML model may not always work well due to user mobility and environmental changes. For example, the user is out of the current model's service area or the surroundings have changed significantly. In such case, a new AI/ML model is required to continue the high-accuracy positioning service for target UEs. The process of model activation and deactivation is needed to flexibly control the model's lifecycle, so as to ensure positioning performance.

When AI/ML model is deployed at UE side and the current model does not work well, network side should send a model deactivation signaling to invalidate the current model. Then, network side may transfer a new model to UE side or instruct UE side to fine-tune the current model. Optionally, falling back to non-AI methods should be also supported. Finally, network side should activate the new model to provide AI/ML based positioning service for UEs. In addition to network-side triggering, UE side can also trigger model activation or deactivation by proactively transferring the model activation or deactivation request to network side. The detailed specification impacts are listed as follows.

- UE side sends model deactivation request to network side when model deactivation is triggered by UE side.

- Network side should send a model deactivation signaling to invalidate the current model.
- Network side may transfer a new model to UE side or instruct UE side to fine-tune the current model.
- UE side sends a model activation request to network side when model activation is triggered by UE side.
- Network side should send a model activation signaling to activate the new model.

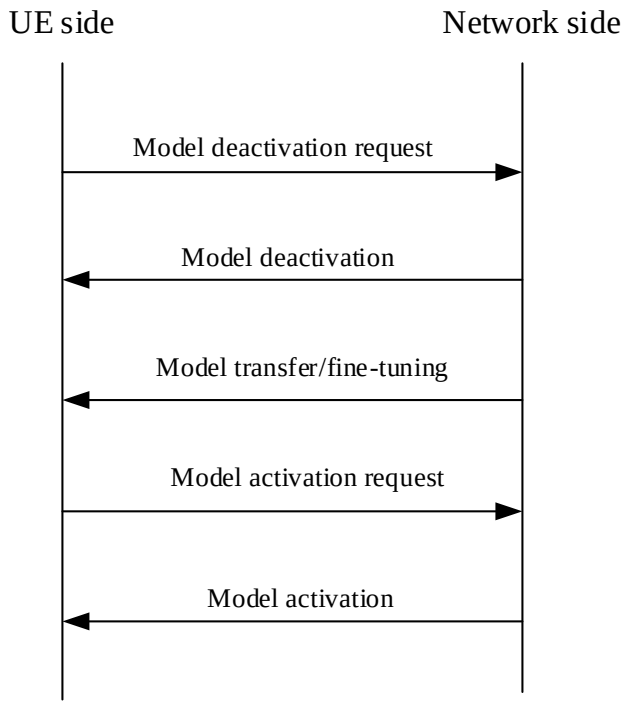

Figure 5 Information exchange procedure for model activation and deactivation

Proposal 12: The process of model activation and deactivation is needed to flexibly control the model's lifecycle, so as to ensure positioning performance.

2.4.3. Model selection

Model selection is the process of selecting a suitable model from a pre-deployed model pool. In practice, considering the dynamics and complexity of the environment, a model pool may be deployed in advance at UE side to enable seamless model switching. When the current model does not work well, network side can indicate the target UE to conduct model selection immediately, so as to adapt to the new environment, which is different from either the aforementioned model transfer or fine-tuning (which is described in detail in section 2.5 as part of model monitoring/updating). In our view, model selection may enable seamless model switching to achieve better service continuity as compared to model transfer or fine-tuning, since the time consumption of parameter transmission over the air-interface and model finetuning can be saved. The detailed specification impacts are listed as follows:

- Network side sends a model pool to UE side in advance, the models in which are used for positioning.
- Network side sends a model selection instruction to the target UE to perform model selection. This instruction may consist of an ID set of candidate models or other assistance information to support UE side perform model selection.
- UE side can select a suitable model from the model pool with reference to the model selection instruction.

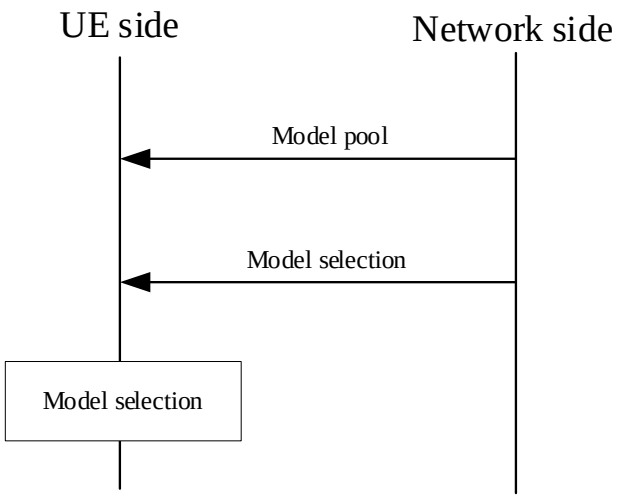

Figure 6 Information exchange procedure for model selection

In particular, to reduce the overhead of model transfer, the deployed model pool can consist of multiple AI/ML models with same structure but different parameters to adapt to varying environments. Meanwhile, each of AI/ML models within the model pool could be associated with a meta-information to assist model selection. For example, each of AI/ML models may be associated with one PRS configuration.

Proposal 13: Network side should deploy a model pool containing multiple models with same structure but different parameters to UE in advance for model selection.

Proposal 14: Network side could send a model selection instruction to instruct the target UE to select a suitable model from the model pool, when the current model does not work well.

2.5. Model monitoring

At the RAN1#111 meeting, it was agreed that:

Agreement

- Regarding AI/ML model monitoring for AI/ML based positioning, to study and provide inputs on feasibility, potential benefits (if any) and potential specification impact at least for the following aspects
  - At least the following are identified for further study as potential data for calculating monitoring metric
    - If monitoring based on model output
      - E.g. , estimated UE location corresponding to model output for direct AI/ML positioning, estimated intermediate parameter(s) corresponding to model output for AI/ML assisted positioning, ground truth label corresponding to model inference output for both direct and AI/ML assisted positioning
    - If monitoring based on model input
      - E.g., measurement corresponding to model inference input
    - Note1: other type of potential data for model monitoring is not precluded
    - Note2: combination of one or more type of potential data for monitoring is not precluded
  - If a given type of data is necessary for calculating monitoring metric, study whether and if so
    - How an entity can be used to provide the given type of data for calculating monitoring metric
      - Companies are requested to report their assumption of the entity (or entities) used to provide the given type of data for calculating monitoring metric for each case
    - Potential signalling for provisioning of the given type of data for calculating associated monitoring metric
    - Potential assistance signaling and procedure to facilitate an entity providing data for calculating monitoring metric
    - Potential UE-network interaction
      - E.g., model monitoring decision indication between UE and network

At the RAN1#112 meeting, it was agreed that:

Agreement

Regarding AI/ML model monitoring for AI/ML based positioning, to study and provide inputs on benefit(s), feasibility, necessity and potential specification impact for the following aspects

- Entity to derive monitoring metric

    - UE at least for Case 1 and 2a (with UE-side model)
        - FFS PRU for Case 1 and 2a
    - gNB at least for Case 3a (with gNB-side model)
        - FFS gNB for Case 3b (with LMF-side model)
    - LMF at least for Case 2b and 3b (with LMF-side model)
    - Note1: companies are requested to report their assumption of entity to calculate monitoring metric if different from above options for each of the agreed cases (Case 1 to Case 3b)
- If model monitoring does not require ground truth label (or its approximation).
    - Monitoring metric, e.g., statistics of measurement, relative displacement, inference output inconsistency, etc.
    - Assistance signaling and procedure, e.g., RS configuration(s) for measurement, measurement statistics as compared to the model input statistics of the training data, etc.
    - report of the calculated metric and/or model monitoring decision
- If model monitoring requires and is provided ground truth label (or its approximation)
    - Monitoring metric, e.g., statistics of the difference between model output and ground truth label, etc.
    - Assistance signaling and procedure, e.g., from LMF to UE/gNB indicating ground truth label and/or measurement, etc.
    - report of the calculated metric and/or model monitoring decision
- Note2: other options (of monitoring methods, monitoring metrics, assistance signaling) are not precluded

In RAN1#112bis-e, the following agreement was reached.

Agreement

Regarding monitoring for AI/ML based positioning, at least the following aspects are identified for further study on benefit(s), feasibility, necessity and potential specification impact for each case (Case 1 to 3b)

- Assistance signaling from LMF to UE/PRU/gNB for UE/gNB-side model monitoring
- Assistance signaling from UE/PRU for network-side model monitoring
- Model monitoring based on provided ground truth label (or its approximation)
    - Monitoring metric: statistics of the difference between model output and provided ground truth label
    - Provisioning of ground truth label and associated label quality
- Model monitoring using at least statistics of measurement(s) without ground truth label
    - Monitoring metric: e.g., statistics of measurement(s) compared to the statistics associated with the training data
    - Note1: the measurement(s) may or may not be the same as model input

Note2: other monitoring methods (e.g., based on statistics of model output without ground truth label, based UE motion sensor and/or jointly based on multiple monitoring metrics) are not precluded

There is no guarantee that the model's performance during its deployment will be consistent with its performance during evaluation. There are various reasons for that, such as training based on non-representative samples, as well as the dynamic and ever-changing nature of the world. For AI/ML based positioning, offline-trained AI/ML model may not always work well due to the nonstationary wireless propagation environments, such as the change of cabinet location in the factory. In such case, we can fine-tune or retrain the model with some newly collected field data, and even activate another AI/ML model directly. But before that, the first step is to make the network (network side and UE side) aware that the current model is no longer valid. Therefore, it is essential to monitor the performance of the model over time so as to ensure the quality of its output on a continuous basis.

2.5.1. Theory analysis for model monitoring

To motivate the model monitoring schemes, we present some general concepts and further analyze the principle of model monitoring from the perspective of Bayesian theory and dataset shift [4]. In particular, we devote to clarify the differences in principle, availability and reliability of model input based model monitoring and model output based model monitoring, so as to lighten future study on this topic.

From the perspective of Bayesian theory, some basic concepts on model monitoring are clarified firstly:

- Dataset shift: the joint distribution of inputs and outputs $\boldsymbol{p}(\boldsymbol{x},\boldsymbol{y})$ differs between training and test stage.
- Covariate (input) shift: only the input distribution $\boldsymbol{p}(\boldsymbol{x})$changes, whereas the conditional distribution of the outputs given the inputs $\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})$ remains unchanged.

- Concept (translation)shift: the conditional distribution of the output given the inputs $\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})$ changes, whereas the input distribution $\boldsymbol{p}(\boldsymbol{x})$remains unchanged.

Generally, covariate shift focuses on the shift of the input distribution $p(x)$, ignoring the condition that $p(y|x)$ remains unchanged. Concept shift focuses on the shift of the conditional distribution $p(y|x)$, ignoring the condition that $p(x)$remains unchanged.

According to Bayesian formula:

$$p(x,y) = p(y|x)p(x)$$

The joint distribution $p(x,y)$ contains the most intrinsic law about dataset generation, depending on the distribution of the inputs $p(x)$ and the mapping that connects the input space to the output space $p(y|x)$. Once the environment changes, it is expected that the joint distribution changes accordingly. In practice, it is difficult to detect the change of the joint distribution (dataset shift) directly, and detecting the changes of $p(x)$ and $p(y|x)$ is more common. When $p(y|x)$ and $p(x)$ change, $p(x,y)$ will change accordingly.

How to connect Bayesian theory with Machine learning theory? There are two kinds of Machine learning frameworks, i.e., generative model (e.g., GAN) and discriminative model (e.g., AI/ML model for positioning). The generative model captures the intrinsic law on data generation, i.e., the joint distribution $p(x,y)$. The discriminative model captures the mapping between the input space to the output space, i.e., the condition distribution $p(y|x)$. In this contribution, we only focus on model monitoring of the discriminative model.

- The first method is to detect the concept shift directly, i.e., the change of mapping between the input space and the output space, which is model output based model monitoring. For example, the positioning results of AI/ML model deviate from the ground truth labels obtained by some known PRUs. Monitoring for concept shift is able to give a definite answer on whether the deployed model is still accurate or not. However, it is impossible to detect concept shift just based on the outputs of the model. Additional information is required that helps to indicate whether the outputs are accurate for inputs, such as ground truth labels.

- The second method is to detect the covariate shift, i.e., the change of $p(x)$, which is model input based model monitoring. From Bayesian formula, the distribution $p(x)$ is not directly related to $p(y|x)$, and the presence of covariate shift does not necessarily invalidate the outputs of the model, but it is certainly a trigger for further validation on model performance in such condition. As a result, it can be used as an indirect indicator of degraded performance since it can be detected directly from the inputs, if additional information about $p(y|x)$ is not easily available. As shown in Figure 7, we summarize three cases for covariate shift.
  - For case (a), the distribution $p(x)$ of training dataset completely covers the distribution of test dataset, but there still exists a sample selection bias between training dataset and test dataset. The presence of case (a) only indicates that the current model may suffer from degraded performance, and the final conclusion on model monitoring still needs further verification.
  - For case (b), there is an overlap between the distributions $p(x)$of training dataset and test dataset. Besides sample selection bias, the model is partially invalid at least for these out-of-distribution samples.
  - For case (c), the distributions of training dataset and test data set are completely different. In such case, the model is completely invalid for test dataset.

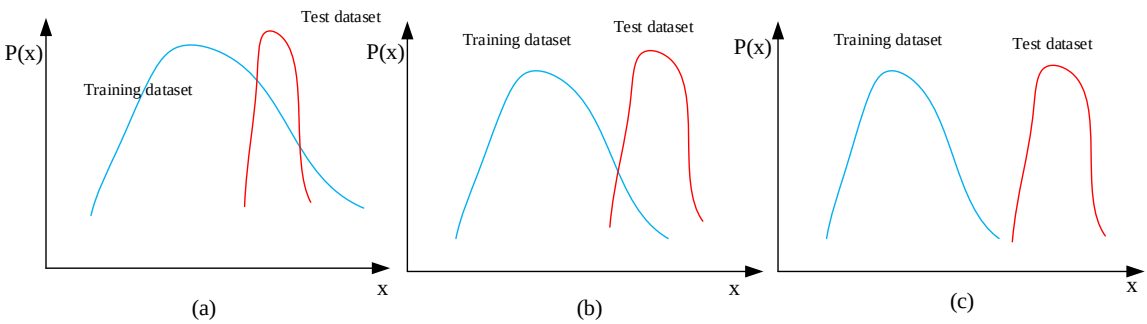

Figure 7 Illustration of three cases for covariate shift

For model input based model monitoring, the basic logic is that we can not expect AI/ML model gives an accurate output for these out-of-distribution (unseen) inputs.

Apart from above two methods, there are also other model monitoring methods from Bayesian formula, such as detecting the shift of $p(y)$ and $p(x|y)$, and their connection to machine learning theory is similar as analyzed above.

$$p(x,y) = p(y|x)p(x) = p(y)p(x|y)$$

From the duration of model failure, there are two granularities of model monitoring:

- **Short-term model monitoring**: The current model only fails at certain locations or at certain times due to some unexpected factors, such as the shading from moving vehicles, but works well at other locations or at other times. In such case, it is not necessary to update the current model, such as fine-tuning, retraining or switching. The short-term performance can be monitored according to the immediate output of the model.
- **Long-term model monitoring**: The current model fails at many locations or at many times due to the change of wireless environments, such as the internal structure of the factory. In such case, model updating is required to adapt the model to the new environment. The long-term performance can be monitored according to the long-term statistics for both model input and output based monitoring.

Proposal 15: Model monitoring can be achieved based on the following categories.

a) Monitor covariate (input) shift: detecting whether or to what extent the distribution of model inputs in the test dataset is consistent with the training dataset.

b) Monitor concept (translation) shift: detecting whether or to what extent the mapping between model inputs and model outputs in the test dataset is consistent with the training dataset.

Observation 2: Monitoring convariate (input) shift can acheve model input based model monitoring, and can be used as an indirect indicator of degraded performance due to its accessibility without the needs of model output related information, such as ground truth labels.

Observation 3: Monitoring concept (translation) shift can acheve model output based model monitoring, and can provide a definite answer on whether the deployed model is still accurate or not, with the needs of model output related information, such as ground truth label.

Proposal 16: Multiple monitoring methods can be integrated together to make a model monitoring decision with the considerations of overhead and accuracy.

Proposal 17: Study specific model monitoring schemes and their specification impacts for both model input based model monitoring and model output based model monitoring with considering the shift of covariate (input) and concept (translation).

2.5.2. General interaction framework for Model monitoring

The classification criteria for self-monitoring and non-self-monitoring methods is whether the entire process of model monitoring is implemeneted on the physical entity where the AI/ML model is deployed without the assistance of other physical entities. Particularly, self-monitoring methods have no specification impact and depend entirely on implemention. In this contribution, we mainly focus on the non-self-monitoring methods, since it is believed that non-self-monitoring methods can provide better model monitoring performance and flexibility for implementation. For example, for the cases where AI/ML models are deployed at UE side, model monitoring at NW side is expected to have better model monitoring performance since NW side has the full knowledge of the AI/ML model and is easier to obtain the monitoring related information of all UEs. It is noted that the following model monitoring methods may also be implemented by a self-monitoring manner.

To calculate the monitoring metric, the reference information is required to compare with the model input or output related information. The reference information may come from source training dataset of the monitored AI/ML model or measurement. Specifically, the reference information comes from the measurement for output based model monitoring methods as illustrated in Figure 8, and the reference information comes from the source training dataset for input based model monitoring methods as illustrated in Figure 9. More details can refer to Section 7.3 of our company's contribution [2]. According to model monitoring metric, a model monitoring decision would be made to indicate whether the current model is valid or not.

According to the functionality involved in the process of model monitoring, model monitoring can be divided into three logic entities:

- Entity of data collection for model monitoring: The entity responsible for collecting data, which is used for model monitoring. For example, the entity can collect ground truth labels for output based model monitoring, and collect CIRs for input based model monitoring.
- Entity of metric calculation for model monitoring: The entity responsible for calculating metrics with reference to the collected data for monitoring. For example, a positioning error acting as a monitoring metric can be calculated by comparing the estimated label of AI/ML model and the collected ground truth label.
- Entity of monitoring decision for model monitoring: The entity responsible for making a decision on whether the monitored model is valid or not with reference to the monitoring metric. If not, the subsequent operations can be triggered, such as deactivation/activation, switching, selection and so on.

Note that these three entities are classified from a functional perspective, and physically they can also be the same entity, different elements of the same entity and completely different entities. A general interaction framework for model monitoring is presented as Figure 10, and the specific mapping between the functionality entities to the physical entities is case-by-case.

To support data collection for model monitoring, dedicated reference signal may be configured. For transmitting the collected data from the entity of data collection to the entity of metric calculation, at least the following options can be considered. The first is to transmit the raw data, such as CIR, SINR and so on. The second is to transmit the parameteric representation of the raw data, such as parameters of a given distribution type. Moreover, the monitoring metrics transmitting from the entity of metric calculation to the entity of monitoring decision can be a statistical or non-statistical parameter, which is also case-by-case. After monitoring decision, a monitoring result along should be delivered from the entity of monitoring decision to the entity of model inference.

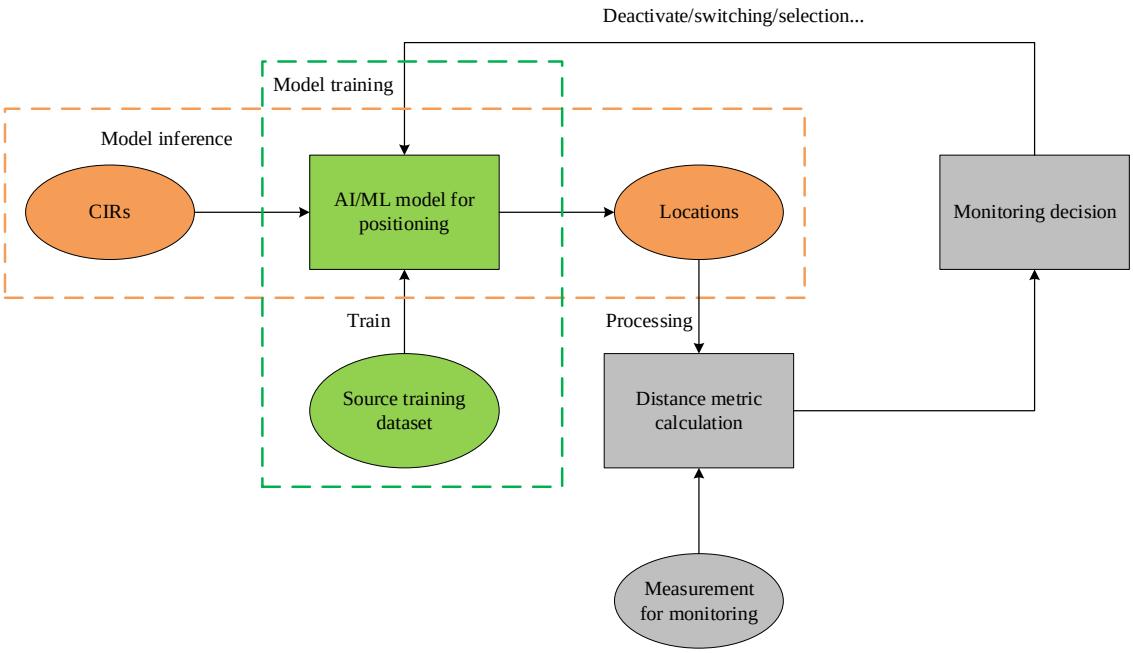

Figure 8 An example of model output based model monitoring

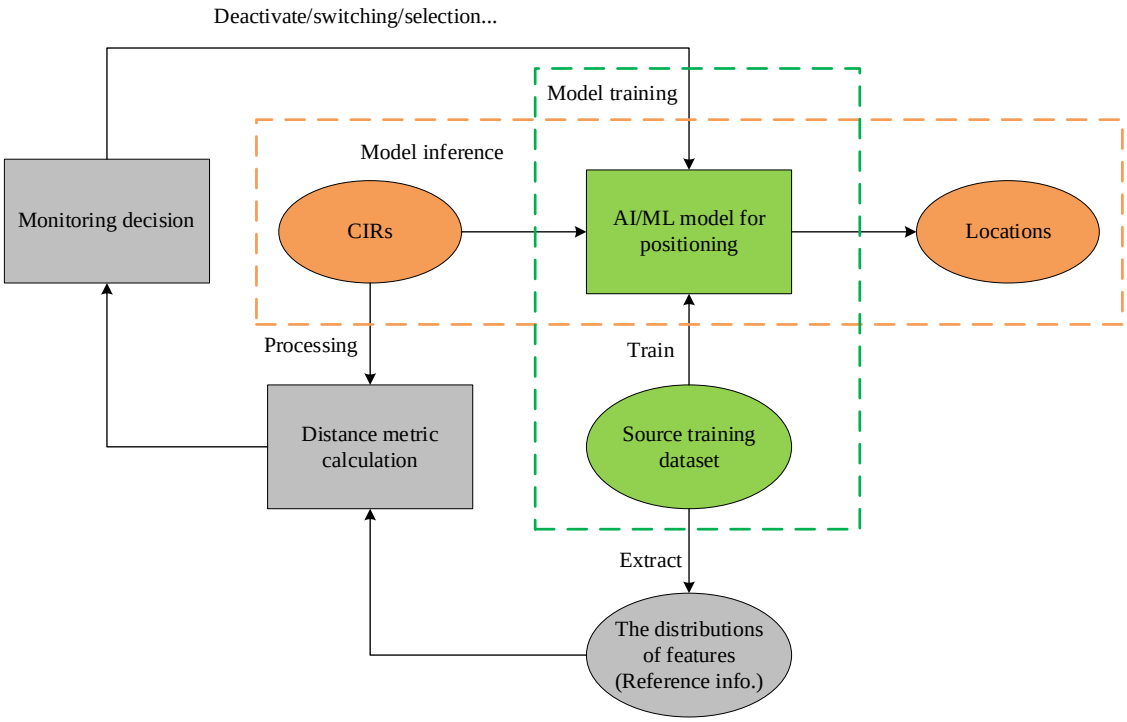

Figure 9 An example of model input based model monitoring

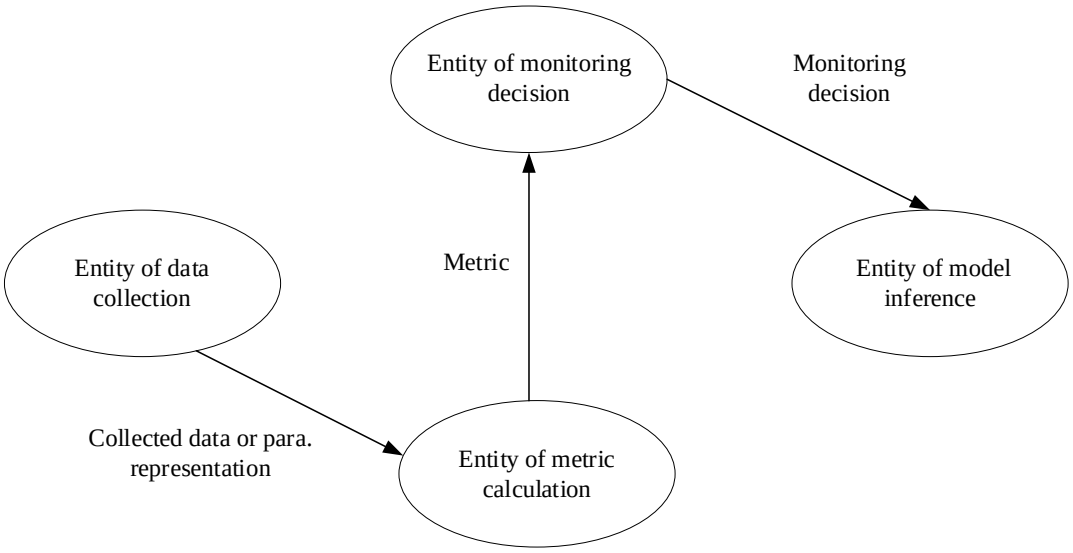

Figure 10 General interaction frametwork for model monitoring

Proposal 18: The classification criteria for self-monitoring and non-self-monitoring methods is whether the entire process of model monitoring is implemented on the physical entity where the AI/ML model is deployed without the assistance of other physical entities.

Proposal 19: According to the functionality involved in the process of model monitoring, model monitoring can be divided into three entities:

- **Entity of data collection for model monitoring**
- **Entity of metric calculation for model monitoring**
- **Entity of monitoring decision for model monitoring**

Proposal 20: For model monitoring, the specific mapping between the logic entities to the physical entities is case-by-case.

Proposal 21: Dedicated reference signal may be configured to support data collection for model monitoring.

2.5.3. Specification impact for model monitoring schemes

In section 2.5.2, the general model monitoring framework is discussed and three logic entities for model monitoring are defined. In our view, the further analysis of specific specification impact on model monitoring should be case-by-case, which is similar to the analysis on legacy positioning methods in pre-Releases. For example, timing based positioning and angle based positioning have different specification impacts in terms of assistance information and metric. Similarly, it is envisioned that model input based model monitoring and model output based model monitoring also have different specification impacts. With reference to the evaluations on model monitoring schemes presented in [2], we further analyze the mapping relationship between the logic entities to the physical entities and the required signalings for different schemes.

Proposal 22: Under the general model monitoring framework, the further analysis of specific specification impact on model monitoring should be case-by-case, including the mapping between the logic entities to the physical entities and the specific signaling.

2.5.3.1. Input-based model monitoring

According to the analysis in section 2.5.1, the model monitoring can base on the covariate (input) shift. Following the idea, we provide the three solutions in sub agenda.

2.5.3.1.1. The shift detection of input

The dominant features, such as first-path delay, RSRP, SINR and delay spread, are relatively easy to measure or estimate from the known RS or CIR. It is expected that the change of these dominant features can affect CIR estimation by a direct or indirect manner, which would further degrade the positioning performance for AI/ML based positioning when CIR is adopted as model input. In this sense, detecting the shift of dominant feature distribution could be an important indicator for model monitoring.

To detecting the shift of dominant feature distribution, there are some issues to resolve firstly:

- **What features are useful for model monitoring?** It is necessary to identify the specific dominant features useful for model monitoring, such as SINR. This conclusion should be dependent on the simulation evaluation.
- **How to detect the shift of dominant feature distribution?** Firstly, there should be a reference distribution of the dominant feature for comparison. The reference distribution as a kind of the reference information is extracted from the source training dataset for model training, which indicates that the AI/ML model works well if the model input obeys this reference distribution. When the distribution of test dataset severely deviates from the reference distribution, it means that the AI/ML model may suffer from degraded performance.
- **What metric is adopted to calculate the difference between two distributions?** Mathematically, there are already many metrics that can describe the difference between two distributions, such as max vertical distance of CDF curve, cross entropy and KL divergence, which can be reused directly here.

When it comes to the specification impact, we have the following proposals.

- When the model training entity and the entity of metric calculation for model monitoring are not at the same side, the reference distribution should be delivered from the model training entity to the entity of metric calculation for model monitoring. Delivering samples of dominant feature is also optional with higher overhead, and the entity of metric calculation can derive the distribution from these samples by some conventional fitting methods, such as GMM.
- When the entity of data collection for model monitoring (measuring the dominant features) and the entity of metric calculation for model monitoring are not at the same side, the measured samples should be delivered from the entity of data collection to the entity of metric calculation. Similarly, delivering the distribution derived from the measured samples is also optional.
- The type of adopted dominant feature for model monitoring should be indicated to the entity of data collection in advance.
- The metric used for calculating the difference between two distributions should be indicated to the entity of metric calculation for model monitoring to align the understanding between the entity of metric calculation and the entity of monitoring decision.
- The model monitoring result should be delivered from the entity of monitoring decision to the entity of model inference.

Adopting Case1 as an example, some possible mappings between the logic entities to the physical entities are listed as Table 2, and other options are not pre-excluded:

Table 2 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	UE	UE
Option 2	NW	UE	UE	UE	NW
Option 3	NW	UE	UE	NW	NW

For option 1 as shown in Figure 11,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), the type(s) of dominant feature(s) for data collection, the reference distribution(s) of dominant feature(s) for metric calculation, the threshold for monitoring decision and other possible information.
- UE side performs model monitoring, including data collection, metric calculation and monitoring decision. After model monitoring, UE side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

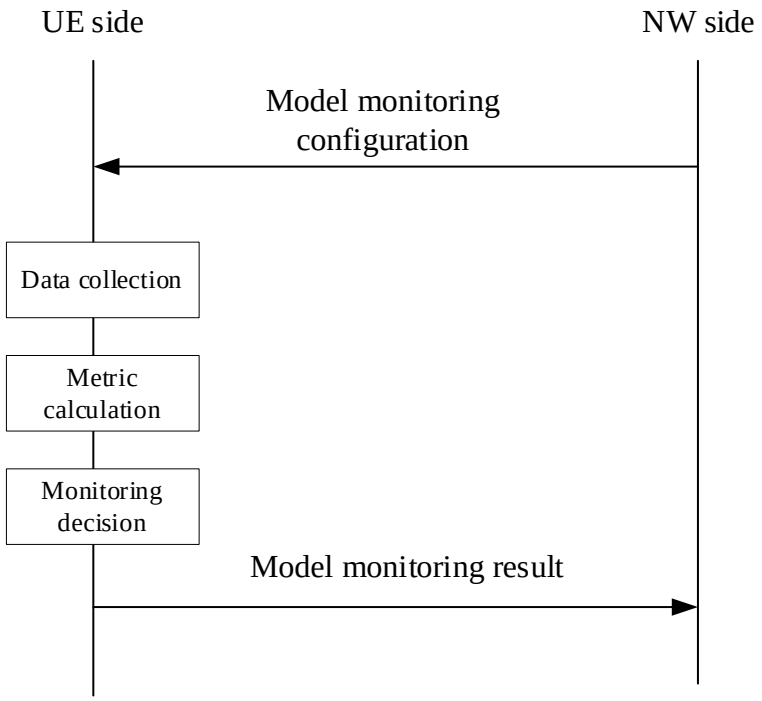

Figure 11 The procedure of model monitoring for option1

For option 2 as shown in Figure 12,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), the type(s) of dominant feature(s) for data collection, the reference distribution(s) of dominant feature(s) for metric calculation and other possible information.
- UE side performs data collection and metric calculation, and then reports the monitoring metric to NW side. The monitoring metric could be the distance between the reference distribution and the distribution of the collected data, such as KL divengence.
- NW side makes monitoring decision with reference to the monitoring metric from UE side. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

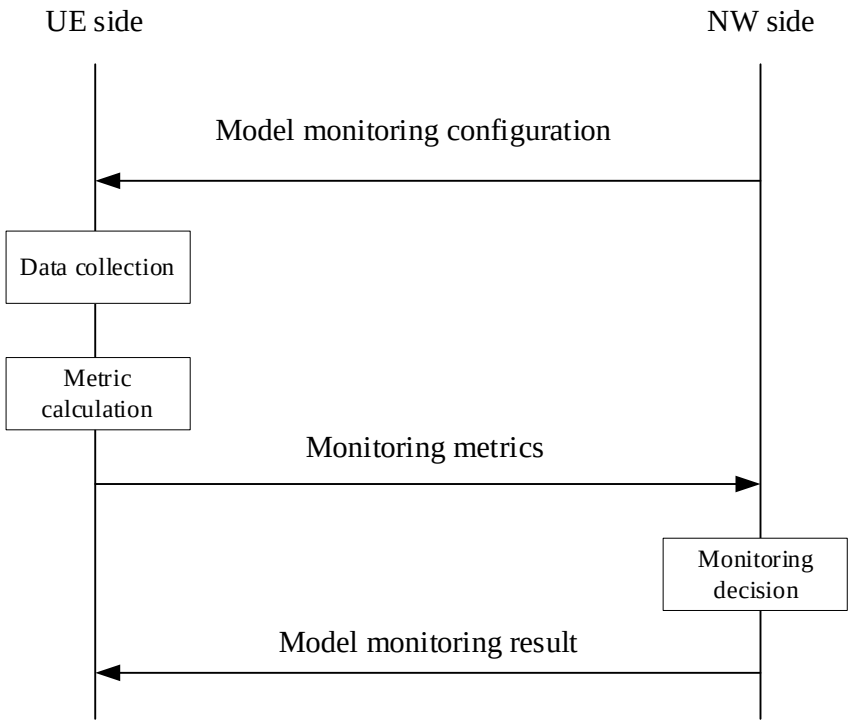

Figure 12 The procedure of model monitoring for option2

For option 3 as shown in Figure 13,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), the type(s) of dominant feature(s) for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. To reduce the overhead, reporting the parameteric representation of the collected data is also optional.
- NW side calculates the monitoring metric with reference to the collected data from UE side.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

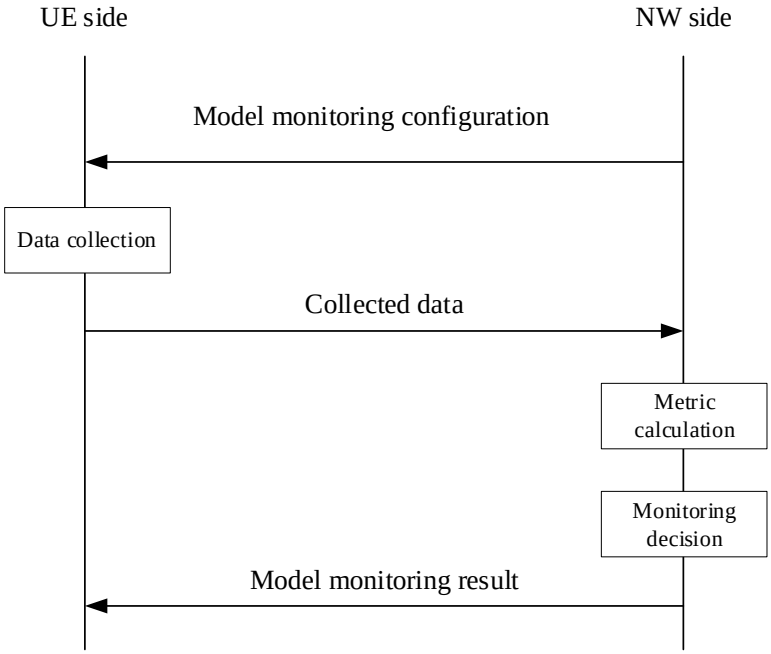

Figure 13 The procedure of model monitoring for option3

Proposal 23: Further study specification impact of the shift detection of dominant feature distribution based model monitoring, may include

a) the type(s) of dominant feature(s) for input data

b) the reference distribution(s) of dominant feature(s)

2.5.3.1.2. Input data classification based on AI/ML adversarial validation

To achieve AI/ML based adversarial validation, the model monitoring entity is required to have the capability of model training, and the whole or part of training data as a kind of reference information is needed to train the classifier. Considering the overheads of model and training data transferred from the model training entity to the model monitoring entity, it is suggested that the model training entity and the model monitoring entity are at the same side, e.g., NW side.

Adopting Case1 as an example, a possible mappings between the functionality entities to the physical entities is listed as Table 3, and other options are not pre-excluded:

Table 3 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	NW	NW

Under this assumption, the specification impacts are listed as follows. The procedure is the same as Figure 13.

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. Note that the collectd data consists of model input related information, such as CIR/PDP.
- NW side trains the classifier based on the collected data from UE side and the source training dataset inside, and then calculates the monitoring metric according to the trained classifier. The monitoring metric could be the accuracy of the trained classifier.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For uplink positioning, there is no specification impact.

Proposal 24: Further study specification impact of AI/ML based adversarial validation based model monitoring, may include

a) The input data collection for monitoring entities for adversarial validation

2.5.3.1.3. Out-of-distribution detection

As for AI/ML based out-of-distribution detection, another AI/ML model is required to monitor the AI/ML model for positioning. Since training the AI/ML model for monitoring requires the whole or part of original training data of the AI/ML model for positioning, it is suggested that these two models should be generated on the same entity such as NW side, to avoid the overhead of large-scale data transfer.

Adopting Case1 as an example, some possible mappings between the functionality entities to the physical entities are listed as Table 4, and other options are not pre-excluded:

Table 4 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	UE	UE
Option 2	NW	UE	UE	UE	NW
Option 3	NW	UE	UE	NW	NW

Under this assumption, the specification impacts are listed as follows.

For Option 1,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s) for metric calculation, the threshold for monitoring decision and other possible information. Note that the reference information refers to the AI/ML model used for monitoring, which stores the distribution information of model input of the original training dataset.
- UE side performs model monitoring, including data collection, metric calculation and monitoring decision. The collected data consists of model input related information, such as CIR/PDP, and the monitoring metric can be the proportion of samples belonging to the original training dataset when given some newly-collected field data. After model monitoring, UE side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

For option2,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s) for metric calculation and other possible information. Note that the reference information refers to the AI/ML model for monitoring, which stores the distribution information of model input of the original training dataset.
- UE side performs data collection and metric calculation, and then reports the monitoring metric to NW side. The collected data consists of model input related information, such as CIR/PDP, and the monitoring metric

can be the proportion of samples belonging to the original training dataset when given some newly-collected field data.

- NW side makes monitoring decision with reference to the monitoring metric from UE side. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For option3,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. The collected data consists of model input related information, such as CIR/PDP.
- NW side inputs the collected data from UE side to the AI/ML model for monitoring, and then calculate the monitoring metric with reference to the output of AI/ML model. The monitoring metric can be the proportion of samples belonging to the original training dataset when given some newly-collected field data.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For uplink positioning, it is better to perform model monitoring at NW side, and there is no specification impact.

Proposal 25: During training the AI/ML model for positioning, a seperate model could be tailored to continually detect whether the model input is out-of-distribution, so as to monitor the performance of the AI/ML model for positioning, with limited resource consumption.

Proposal 26: Further study specification impact of AI/ML based out-of-distribution based model monitoring, may include:

a) AI/ML model for monitoring tranfered from NW side to UE side.

2.5.3.2. Output-based model monitoring

Based on the analysis, model monitoring can divide into short term and long term model monitoring. And considering positioning should be contiguous movement, we provide the following solutions based on the one-shot positioning result or motion-based monitoring solution

2.5.3.2.1. Ground truth label based model monitoring

As for ground truth label based model monitoring, the main specification impact mainly focus on the procedure of collecting the samples with ground truth labels, which can be divided into the scope of data collection.

Adopting Case1 as an example, some possible mappings between the functionality entities to the physical entities are listed as Table 5, and other options are not pre-excluded:

Table 5 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	PRU	PRU	PRU	PRU
Option 2	NW	PRU	PRU	PRU	NW
Option 3	NW	PRU	PRU	NW	NW

Under this assumption, the specification impacts are listed as follows. Particularly, the reference information is the measurement of ground truth lables.

For Option 1,

- NW side should indicate related model monitoring configuration as assistance information to PRU firstly, which may consist of model monitoring method(s), indication for data collection, the threshold for monitoring decision and other possible information.

- PRU performs model monitoring, including data collection, metric calculation and monitoring decision. The collected data consists of CIRs for model inference and the related ground truth labels. The monitoring metric could be the positioning error of 90% CDF. After model monitoring, PRU side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

For option2,

- NW side should indicate related model monitoring configuration as assistance information to PRU firstly, which may consist of model monitoring method(s), indication for data collection, and other possible information.
- PRU side performs data collection and metric calculation, and then reports the monitoring metric to NW side. The collected data consists of the estimated positions and the related ground truth labels. The monitoring metric can be the positioning error of 90% CDF.
- NW side makes monitoring decision with reference to the monitoring metric from PRU. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For option3,

- NW side should indicate related model monitoring configuration as assistance information to PRU firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- PRU side performs data collection, and then reports the colleted data to NW side. Note that the collected data consists of the estimated positions and the related ground truth labels.
- NW side calculates the monitoring metric according to the collected data from PRU. The monitoring metric can be the positioning error of 90% CDF.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

Proposal 27: The main specification impact of ground truth label based model monitoring is the procedure of collecting the samples with ground truth labels, which could reuse that of data collection for training.

2.5.3.2.2. Monitoring for AI/ML assisted positioning

To support monitoring for AI/ML assisted positioning, the locations of TRPs are required as an assistance information, whose transfer from NW side to UE side has been supported by pre-releases. For three constructions of AI/ML assisted positioning , this scheme is more suitable for case 2 (single TRP, same model for N TRPs) and case 3 (multi-TRP, one model for N TRPs) since only one model needs to monitored. For case 1 (single-TRP, N models for N TRPs), this scheme can indicate that at least one model is out-of-work among multiple selected models for TOA estimation. It is noted that aforementioned model monitoring methods can also be used to monitor the models for AI/ML assisted positioning.

Adopting Case2a) as an example, some possible mappings between the functionality entities to the physical entities are listed as Table 6, and other options are not pre-excluded:

Table 6 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	UE	UE
Option 2	NW	UE	UE	UE	NW
Option 3	NW	UE	UE	NW	NW

The specific specification impacts are listed as follows.

For Option 1,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s)

for metric calculation, the threshold for monitoring decision and other possible information. Note that the reference information refers to the reference distribution when the current AI/ML model is valid.

- UE side performs model monitoring, including data collection, metric calculation and monitoring decision. After model monitoring, UE side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

For option2,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s) for metric calculation and other possible information. Note that the reference information refers to the reference distribution when the current AI/ML model is valid.
- UE side performs data collection and metric calculation, and then reports the monitoring metric to NW side. Note that the collected data consists of the estimated TOAs related to multiple TRPs, and the monitoring metric is calculated according to the locations of TRPs and related TOA estimations.
- NW side makes monitoring decision with reference to the monitoring metric from UE side. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For option3,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. Note that the collected data consists of the estimated TOAs related to multiple TRPs.
- NW side calculates the monitoring metric according to the estimated TOAs from UE side and TRPs' location.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

Proposal 28: Further study specification solution and impact on model monitoring for AI/ML assisted positioning, may include

a) The error threshold between AI output and estimated timing based on calculated location

b) The timing measurement enhancement indication based AI/ML assisted positioning

2.5.3.2.3. Motion sensors assisted model monitoring

As listed in Table 7, release 17 has supported the report of sensor measurement information from UE to the LMF [5], and these information can be utilized to assist model monitoring at NW side. From the perspective of UE privacy, model monitoring can also be performed at UE side without need of reporting sensor measurement information if UE does not want to disclose their privacy information to NW side.

Table 7 Sensor Measurement Information that may be transferred from UE to the LMF [5]

Information	UE-assisted	UE-based/ Standalone
Displacement Timestamp	Yes	Yes
Displacement Information	Yes	Yes
Reference Position	Yes	Yes
Reference Time	Yes	Yes

Adopting Case1 as an example, some possible mappings between the functionality entities to the physical entities are listed as Table 8, and other options are not pre-excluded:

Table 8 Mapping between the logic entities to the physical entities

Items			The entity of model monitoring

	The entity of model training	The entity of model inference	The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	UE	UE
Option 2	NW	UE	UE	UE	NW
Option 3	NW	UE	UE	NW	NW

The specific specification impacts are listed as follows. The reference information is the measurement of motion sensors.

For Option 1,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the threshold for monitoring decision and other possible information.
- UE side performs model monitoring, including data collection, metric calculation and monitoring decision. The collected data consists of the CIR/PDP for model inference and motion sensor information for displacement estimation. The metric can be the distance between the distributions of the displacement estimated by AI/ML model and displacement estimated by motion sensors. After model monitoring, UE side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

For option2,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection and metric calculation, and then reports the monitoring metrics to NW side. Note that the collected data consists of the CIRs for model inference and motion sensor information for displacement estimation. The metric can be the distance between the distributions of the displacement estimated by AI/ML model and displacement estimated by motion sensors
- NW side makes monitoring decision with reference to the monitoring metric from UE side. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For option3,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. Note that the collected data consists of the CIRs for model inference and motion sensor information for displacement estimation.
- NW side calculates the monitoring metric according to the collected data from UE side. The metric can be the distance between the distributions of the displacement estimated by AI/ML model and displacement estimated by motion sensors
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

Proposal 29: Further study specification impact of motion sensors assisted model monitoring considering the UE privacy, may include:

a) Indication to inform UE side to collect motion sensor information.

2.5.3.2.4. Ranging model assisted model monitoring

This method monitors the positioning model via another AI/ML model called ranging model. Both the positioning model and the ranging model are generated using the same training dataset. However, the ranging model produces the distance between two CIRs as its output. It is said that the positioning model captures the mapping relationship between the input space to the physical space, while the ranging model captures the mapping relationship between the input space to the latent space. Therefore, to avoid the overheads of ranging model or training data transferred

from the model training entity to the model monitoring entity, it is suggested that the model training entity and the model monitoring entity are at the same side, e.g., NW side.

Adopting Case1 as an example, a possible mappings between the functionality entities to the physical entities is listed as Table 9, and other options are not pre-excluded:

Table 9 Mapping between the logic entities to the physical entities

Items	The entity of model training	The entity of model inference	The entity of model monitoring		
			The entity of data collection	The entity of metric calculation	The entity of monitoring decision
Option 1	NW	UE	UE	UE	UE
Option 2	NW	UE	UE	UE	NW
Option 3	NW	UE	UE	NW	NW

Under this assumption, the specification impacts are listed as follows. Note that the reference information refers to the ranging model, which stores the mapping relationship between the input space and the latent space. Given two inputs, the ranging model can output their distance in the latent space.

For Option 1,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s) for metric calculation, the threshold for monitoring decision and other possible information.
- UE side performs model monitoring, including data collection, metric calculation and monitoring decision. The collected data consists of model input related information, such as CIR/PDP, and the monitoring metric can be the distance between the distributions of the distance in the physical space and the distance in the latent space. After model monitoring, UE side should report the model monitoring result to NW side for further operations, such as activation/deactivation, switching and updating.

For option2,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection, the reference information(s) for metric calculation and other possible information. Note that the reference information refers to the AI/ML model for monitoring, which stores the distribution information of model input of the original training dataset.
- UE side performs data collection and metric calculation, and then reports the monitoring metric to NW side. The collected data consists of model input related information, such as CIR/PDP, and the monitoring metric can be the distance between the distributions of the distance in the physical space and the distance in the latent space.
- NW side makes monitoring decision with reference to the monitoring metric from UE side. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For option3,

- NW side should indicate related model monitoring configuration as assistance information to UE side firstly, which may consist of model monitoring method(s), indication for data collection and other possible information.
- UE side performs data collection, and then reports the colleted data to NW side. The collected data consists of model input related information, such as CIR/PDP.
- NW side inputs the collected data from UE side to the AI/ML model for monitoring, and then calculate the monitoring metric with reference to the output of AI/ML model. The monitoring metric can be the distance between the distributions of the distance in the physical space and the distance in the latent space.
- NW side makes monitoring decision with reference to the monitoring metric. Then, NW side may informs the monitoring result along with other indications, such as activation/deactivation, switching and updating, to UE side.

For uplink positioning, it is better to perform model monitoring at NW side, and there is no specification impact.

Proposal 30: During training the AI/ML model for positioning, a seperate ranging model could be tailored to continually monitor the performance of the AI/ML model for positioning, with limited resource consumption.

Observation 4: With the assistance of the ranging model, efficient model monitoring can be achieved without the needs of other measurement related to model output.

Proposal 31: Further study specification impact of ranging model assisted model monitoring, may include:

a) Ranging model transferred from NW side to UE side for monitoring.

2.6. Model updating/fine-tuning

In our companion contribution [2], we have shown that model fine-tuning can significantly improve model generalization performance. To adapt to the dynamic wireless environment quickly, model fine-tuning is necessary.

In general, model fine-tuning can be regarded as model updating with prior model information and a small scale field data. In this section, we further analyze the detailed specification impacts when model fine-tuning is included.

For the case in which AI model inference is at UE side, fine-tuning can be conducted at UE side based on the pre-trained model. The detailed specification impacts are listed as follows.

- UE side needs to inform the network of device capability in advance. The device capabilities may consist of computation capability, available computation resource, available storage resource, battery status, data collection, and other capabilities related to model fine-tuning.
- Based on the received device capability from the target UE, if network side identifies that the device can support model fine-tuning, network side delivers assistance information including pre-trained model information and training configuration to the target UE to support model fine-tuning. The model information may consist of the input and output of the model, the architecture of the model, the weight of the model, the configuration of the model, the state of the optimizer, and so on. The training configuration consists of the learning rate, training epoch, batch size, end condition, and so on.

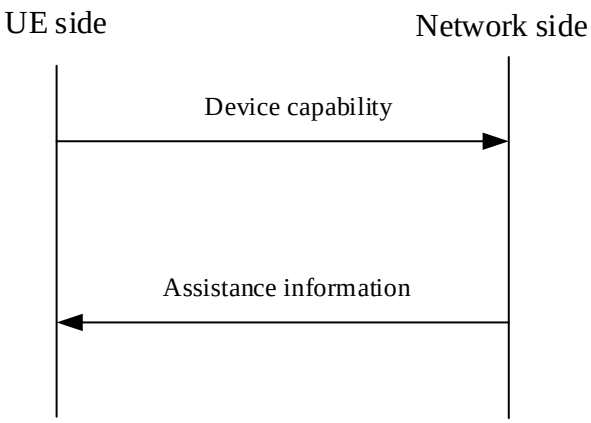

Figure 14 Information exchange procedure when fine-tuning is conducted at UE side

Proposal 32: When fine-tuning is conducted at UE side, UE capability corresponding to fine-tuning is required.

Proposal 33: To enable model fine-tuning when AI/ML model inference is at UE side, support assistance information to the target UE about pre-trained model and training configuration.

For the case that AI model inference is at UE or network side, fine-tuning can be conducted at network side based on the pre-trained model. In this case, UE side does not need to inform network side of device capability, and network side does not need to transfer model information and training configuration to the target UE. To support model fine-tuning at network side, UE should feed back some requested training data to network side. The detailed specification impacts are listed as follows.

- Network side should send data collection requests to the target UEs in advance. The data collection request may consist of data format, privacy permission, training data collection configuration, training data reporting configuration, and so on.
- UE should provide training data to network side to support model fine-tuning if available.

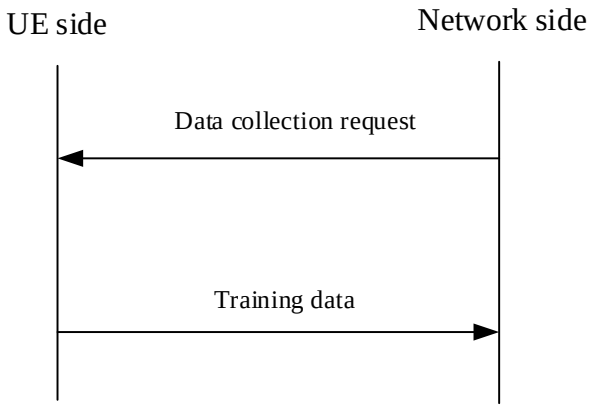

Figure 15 Information exchange procedure when fine-tuning is conducted at network side

Proposal 34: Training data collection request for model fine-tuning and feedback from the target UE is required to support model fine-tuning at network side.

Regarding the occasion or condition for triggering a model updating process, there are two points to consider:

- The current model cannot work well. For example, the result of model monitoring indicates that the current model is out of work.
- Model updating operation can be supported in terms of available training data and hardware resource.

Proposal 35: The result of model monitoring and the achievability of model updating should be jointly considered as the condition of model updating.

2.7. Model inference

At the RAN1#111 meeting, it was agreed that:

Agreement

Regarding AI/ML model inference, to study and provide inputs on potential specification impact (including necessity and applicability of specifying AI/ML model input and/or output) at least for the following aspects for each of the agreed cases (Case 1 to Case 3b) in AI/ML based positioning accuracy enhancement

- Types of measurement as model inference input
  - new measurement
  - existing measurement
- UE is assumed to perform measurement as model inference input for Case 1, Case 2a and Case 2b; TRP is assumed to perform measurement as model inference input for Case 3a and Case 3b
  - Report of measurements as model inference input to LMF for LMF-side model (Case 2b and Case 3b)
- For AI/ML assisted positioning, new measurement report and/or potential enhancement of existing measurement report as model output to LMF for UE-assisted (Case 2a) and NG-RAN node assisted positioning (Case 3a)
- Assistance signaling and procedure to facilitate model inference for both UE-side and Network-side model
  - New and/or enhancement to existing assistance signaling
  - Note: whether such assistance signaling and procedure can be applied to other aspect(s) of AI/ML model LCM can also be discussed

At the RAN1#112 meeting, it was agreed that:

Agreement

Regarding AI/ML model inference, to study the potential specification impact (including the feasibility, and the necessity of specifying AI/ML model input and/or output) at least for the following aspects for AI/ML based positioning accuracy enhancement

- For direct AI/ML positioning (Case 2b and 3b), type of measurement(s) as model inference input considering performance impact and associated signaling overhead
  - Potential new measurement: CIR/PDP
  - existing measurement: e.g., RSRP/RSRPP/RSTD
  - Note1: details of potential new measurement and/or potential enhancement to existing measurement is to be studied

    - Note2: study the impact of model input for other cases are not precluded
- For AI/ML assisted positioning with UE-assisted (Case 2a) and NG-RAN node assisted positioning (Case 3a), measurement report to carry model output to LMF
    - new measurement report: e.g., ToA, path phase
    - existing measurement report: e.g., RSTD, LOS/NLOS indicator, RSRPP
    - enhancement of existing measurement report: e.g., soft information/high resolution of RSTD
- Assistance signaling and procedure to facilitate model inference for both UE-side and Network-side model
    - RS configurations
    - Other assistance information is not precluded

From the perspective of model deployment, this section discusses the detailed specification impacts of AI/ML model inference for direct AI/ML positioning and AI/ML assisted positioning respectively. In our companion contribution [2], the simulation results confirm that adopting time domain CIR/PDP as model input can achieve significantly better positioning accuracy as compared with adopting legacy RSRPP as model input. Moreover, the reporting overhead of CIR/PDP can be greatly reduced by using CIR pre-processing and compression technologies, such as truncation, quantizaion, compression and so on, and is comparable to that of legacy RSRPP. Therefore, it is suggested to support time-domain CIR/PDP as model input for AI/ML based positioning.

Proposal 36: Support time domain CIR/PDP as model input for AI/ML based positioning.

2.7.1. For direct AI/ML positioning

When the AI/ML model has already been deployed and activated at UE side, UE side can perform positioning using this model. For direct AI/ML positioning, there is no extra specification impart when model inference is performed at UE side.

In the case of AI/ML model inference at the network side, UE side does not need to inform network side of device capability, and network side does not need to deliver model information to the target UE. To enable direct AI/ML positioning, information exchange is still necessary. The detailed specification impacts are listed as follows.

- Network side should inform UE side of the required measurement type (e.g., CIR as input to AI/ML model for inference).
- UE side reports requested measurement feedback.
- Network side performs position estimation and then transfers positioning results to the target UE if the positioning application is originated from UE side.

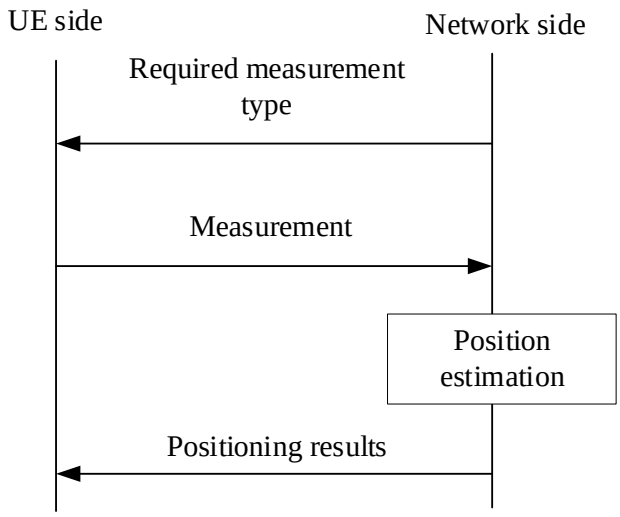

Figure 16 Information exchange procedure for direct AI/ML positioning when model inference is at network side

Proposal 37: For direct AI/ML positioning, when model inference is at network side, request to and feedback from the target UE of the necessary measurement (e.g., as the input to the AI/ML model) for model inference is needed.

2.7.2. For AI/ML assisted positioning

When the AI/ML model has already been deployed and activated at UE side, UE side can utilize this AI/ML model to extract the intermediate feature. The detailed specification impacts are listed as follows.

- UE side reports the intermediate feature to network side where the intermediate feature is extracted by the AI/ML model.
- Network side acting as an Actor performs position estimation according to the received intermediate feature, and then feeds back the positioning results to the target UE if the positioning application originates from UE side.

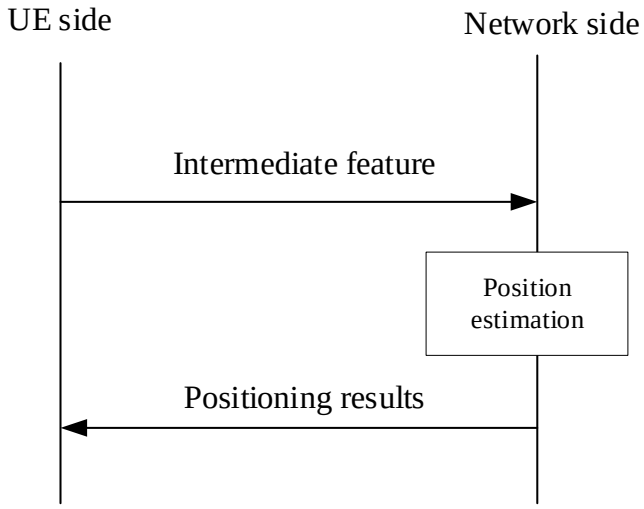

Figure 17 Information exchange procedure for AI/ML assisted positioning when model inference is at UE side

Proposal 38: For AI/ML assisted & UE assisted positioning, support the target UE to report the output of AI/ML model inference (intermediate feature for positioning) when model inference is at UE side.

In the case of AI/ML assisted positioning and AI/ML model inference is at the network side, UE side does not need to inform network side of device capability, and network side does not need to deliver model information to the target UE. To enable AI/ML assisted positioning, the interaction of assistance information is still necessary. The detailed specification impacts are listed as follows.

- Network side should inform UE side of the required measurement type (e.g., CIR as input to AI/ML model for inference).
- UE side reports requested measurement.
- Network side utilizes the AI/ML model to extract intermediate features from received measurement, and performs position estimation by conventional positioning methods. Then, networks side sends the positioning results to the target UE if the positioning application originates from UE side.

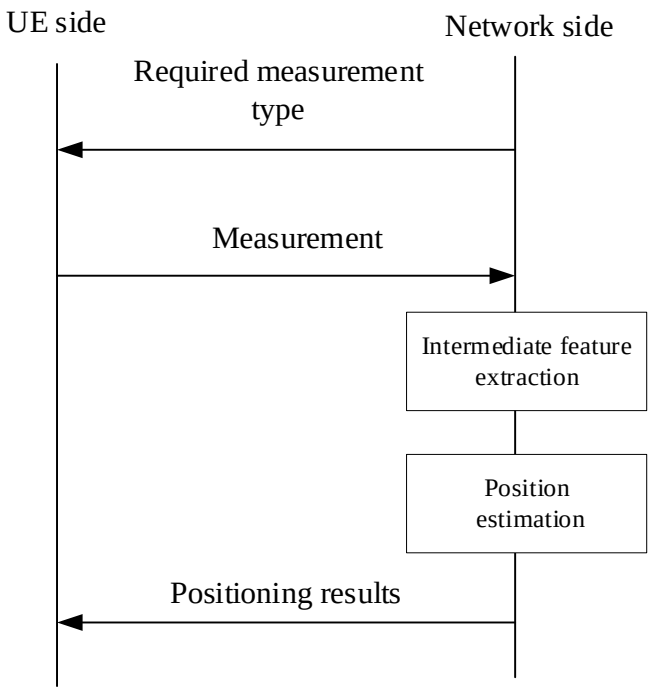

Figure 18 Information exchange procedure for AI/ML assisted positioning when model inference is at network side

In our companion contribution [2], it is verified that incorporating TRP-related information into model input for model inference can significantly improve the inference accuracy for Construction 2 (Single model, same TRP for N model) of AI/ML assisted positioning without additional signaling overhead. For various cases, it is meaningful to identify the performance gain and signaling overhead brought by additional information. At least, TRP-related information, such as encoded TRP ID and TPR's location, should be incorporated into the model input for Construction 2 of AI/ML assisted positioning.

For three constructions of AI/ML assisted positioning, we have the following considerations for perspective of overhead of model deployment and inference.

- For Construction 1 (single TRP, N model for N TRPs), uplink positioning should be primarily considered, and AI/ML model should be deployed at gNB side, i.e., different TRPs deploy their own models. For uplink positioning, gNB side estimates uplink CIRs from uplink SRS, and TOAs associated with multiple TRPs can be estimated when CIRs are input into AI/ML models. Then, gNB side delivers the estimated TOAs to LMF for location calculation.
- For Construction 2 (single TRP, same model for N TRPs), downlink positioning should be primarily considered, and AI/ML model should be deployed at UE side. When performing model inference, TRP-related information should be incorporated into model input, such as encoded TRP's ID, TRP's location.
- For Construction 3 (Multi-TRP, one model for N TRPs), this case is similar to direct AI/ML positioning. Uplink positioning should be primarily considered, and AI/ML model should be deployed at LMF side.

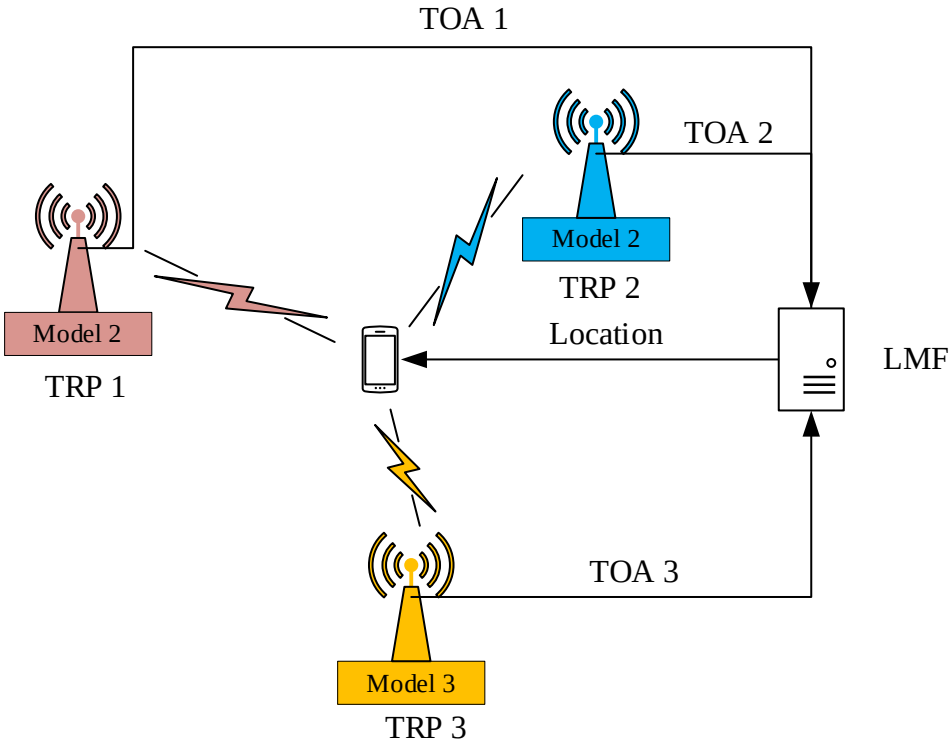

Figure 19 Illustration of Construction 1 (single TRP, N model for N TRPs)

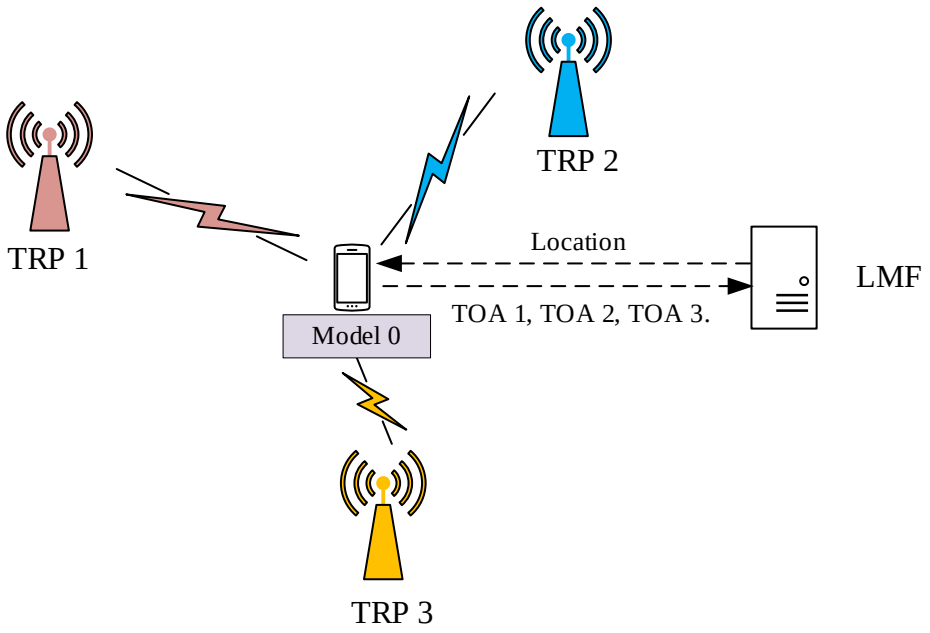

Figure 20 Illustration of Construction 2 (single TRP, same model for N TRPs)

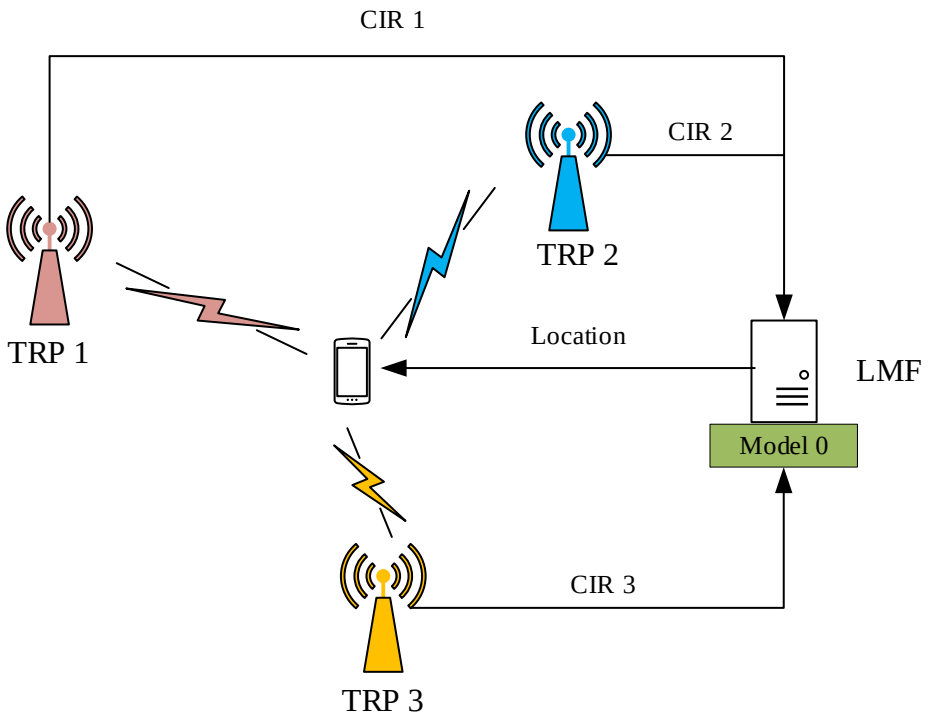

Figure 21 Illustration of Construction 3 (multi-TRP, one model for N TRPs)

From the simulation results and analysis of our company's contribution [2], it is confirmed that TOA based positioning can reap better positioning accuracy as compared with RSTD based positioning since TOA contains more raw information related to positioning, which provides more flexibility for positioning implementation. Therefore, it is suggested to support TOA as a new measurement reporting from UE side to NW side.

Proposal 39: For AI/ML assisted positioning, when model inference is at network side, request to and feedback from the target UE of the necessary measurement (e.g., as the input to the AI/ML model) for model inference is needed.

Proposal 40: TRP-related information, such as encoded TRP ID and TPR's location, should be incorporated into the model input for Construction 2 (Single model, same TRP for N model) of AI/ML assisted positioning without additional signaling overhead.

Proposal 41: For Construction 1 (single TRP, N model for N TRPs), uplink positioning should be primarily considered, and AI/ML models should be deployed at gNB side.

Proposal 42: For Construction 2 (single TRP, same model for N TRPs), downlink positioning should be primarily considered, and AI/ML model should be deployed at UE side.

Proposal 43: For Construction 3 (Multi-TRP, one model for N TRPs), uplink positioning should be primarily considered, and AI/ML model should be deployed at LMF side.

Proposal 44: Support TOA as a new measurement reporting from UE side to NW side.

2.8. A general model management procedure

In the above sections, we have listed common model operations and analyzed their potential specification impacts, respectively. Here, a general model management procedure is presented by integrating all the above model operations. For convenience, only UE based direct AI/ML positioning is given, and similar procedures can be easily extended to other cases.

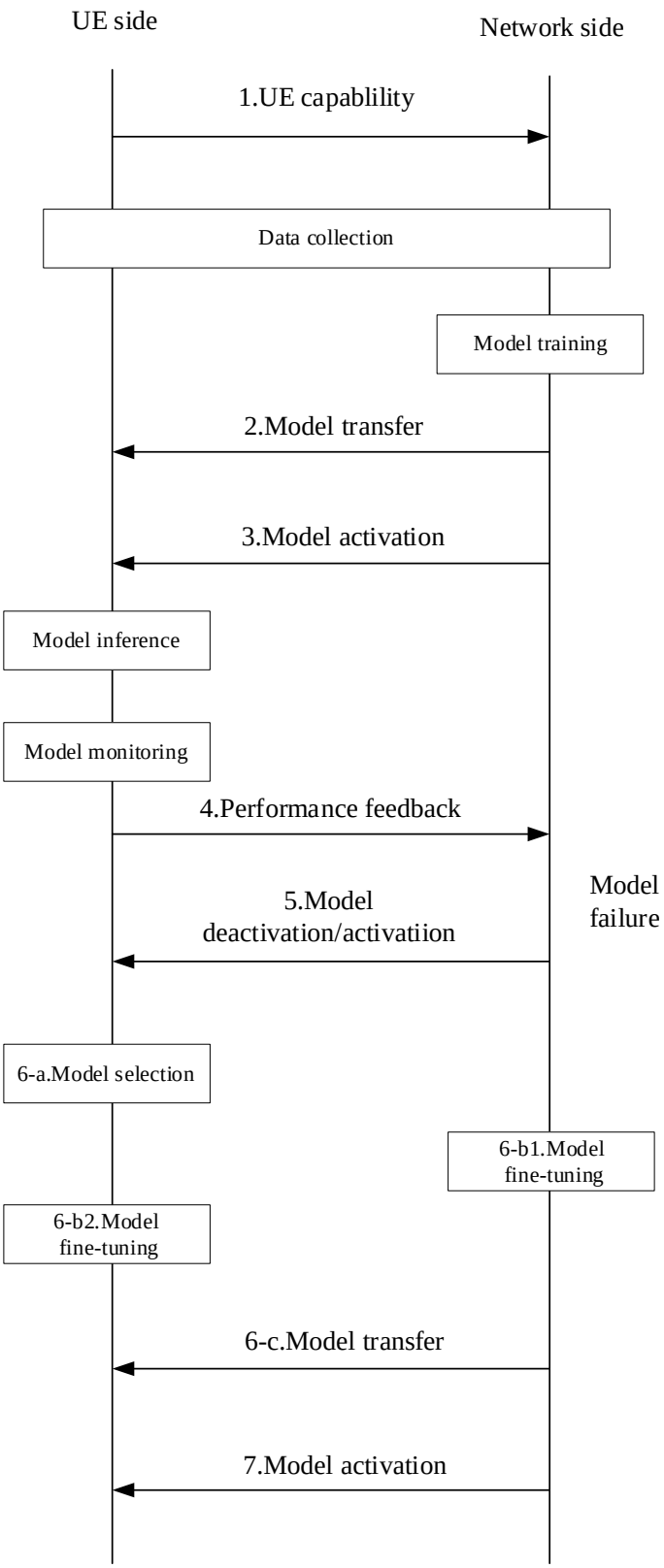

Figure 22 A general model management procedure for UE based direct AI/ML positioning

Proposal 45: A general model management procedure should be specially studied for AI/ML based positioning accuracy enhancement.

2.9. A detailed list of assistance signaling

From section 3.2 to section 3.6, we have analyzed the potential specification impacts of each aspect of model operation in high level, and presented a general model management procedure. There are some requirements for assistance signaling involved in this procedure to support model management. In this section, we further analyze the possible assistance signaling configuration and present a detailed list of assistance signaling.

Table 10 A detailed list of assistance signaling for model training at NW

Procedure of model management		Assistance signaling	
		From network to UE	**From UE to network**
UE capability		/	- Device capability to support model management, e.g., data collection, model training, model inference
Model training	**Data collection**	- PRS/SRS configuration (such as RS pattern, RS resource set) - Data requirement (such as format, type, size, preprocessing, labeling)	- Data collection report, including model input and label
Model indication/configuration	**Model transfer**	- Model information (such as structure, parameters, input type and shape, output type and shape) - Meta information	/
	Model activation/ deactivation	- Dedicated signaling for model activation/deactivation - Model ID to be activated/deactivated	/
	Model selection	- A model pool for positioning - A set of selected model ID	/
Model monitoring and updating	**Model monitoring at network**	- Performance metrics fed back from UE (such as positioning error, channel condition, movement status) - Model validity identification	- Required Performance metrics - Other information useful for model monitoring (such as PRS quality, sensor information, movement status, and positioning results from other positioning methods)
	Model monitoring at UE	- Performance metrics monitored at UE (such as positioning error, channel condition, movement status)	- Model validity identification
	Model updating/ fine-tuning at network side	/	/
	Model updating/ fine-tuning at UE	- Training configuration (such as training method, learning rate, optimizer, epochs) - Pre-trained Model information (such as structure, parameters, input type and shape, output type and shape) - Training data (e.g., collected by UEs)	Model training termination instruction
Model inference at UE	**Direct AI/ML positioning**	/	/

	AI/ML assisted positioning	Required measurement result type(such as ToA, LOS/NLOS identification)	- Intermediate features (such as TOA, LOS/NLOS identification)
Model inference at network	**Direct AI/ML positioning**	- Required measurement type (such as CIR)	- Measurement (such as CIR)
	AI/ML assisted positioning	- Required measurement type (such as CIR)	- Measurement (such as CIR)

Proposal 46: Support to study detailed assistance signaling configuration when the model management procedure for AI/ML based positioning is agreed.

2.10. Discussion on model ID and functionality ID based LCM

At the RAN1 #112bie e-meeting, some agreements on model ID and functionality ID based LCM have been reached in both 9.2.1 and 9.2.4 agendas.

Agreement

- For AI/ML functionality identification and functionality-based LCM of UE-side models and/or UE-part of two-sided models:
    - Functionality refers to an AI/ML-enabled Feature/FG enabled by configuration(s), where configuration(s) is(are) supported based on conditions indicated by UE capability.
    - Correspondingly, functionality-based LCM operates based on, at least, one configuration of AI/ML-enabled Feature/FG or specific configurations of an AI/ML-enabled Feature/FG.
        - FFS: Signaling to support functionality-based LCM operations, e.g., to activate/deactivate/fallback/switch AI/ML functionalities
        - FFS: Whether/how to address additional conditions (e.g., scenarios, sites, and datasets) to aid UE-side transparent model operations (without model identification) at the Functionality level
        - FFS: Other aspects that may constitute Functionality
    - FFS: which aspects should be specified as conditions of a Feature/FG available for functionality will be discussed in each sub-use-case agenda.
- For AI/ML model identification and model-ID-based LCM of UE-side models and/or UE-part of two-sided models:
    - model-ID-based LCM operates based on identified models, where a model may be associated with specific configurations/conditions associated with UE capability of an AI/ML-enabled Feature/FG and additional conditions (e.g., scenarios, sites, and datasets) as determined/identified between UE-side and NW-side.
    - FFS: Which aspects should be considered as additional conditions, and how to include them into model description information during model identification will be discussed in each sub-use-case agenda.
    - FFS: Relationship between functionality and model, e.g., whether a model may be identified referring to functionality(s).
    - FFS: relationship between functionality-based LCM and model-ID-based LCM

Note: Applicability of functionality-based LCM and model-ID-based LCM is a separate discussion

Agreement

Regarding LCM of AI/ML based positioning accuracy enhancement, at least for Case 1 and Case 2a (model is at UE-side), further study the following aspects on information related to the conditions

- What are the conditions for functionality-based LCM
  - which aspects should be specified as conditions of a Feature/FG available for functionality
- What are the conditions for model-ID-based LCM
  - Which aspects should be considered as additional conditions, and how to include them into model description information during model identification

In our understanding, functionality based LCM is a coarse type of management, and only static conditions should be included, such as UE capability related conditions. On the contrary, model-ID based LCM is a fine type of management, and all these dynamic conditions should be included, such as generalization related conditions. Both LCM frameworks could coexist and be integrated to acheve flexbile LCM of AI/ML model.

For functionality based LCM of AI/ML based positioning enhancement, in addition to the existing Feature/FG specified for positioning, it is suggested to focus on AI/ML related Feature/FG. For example, UE side could report model input and output related conditions, such as the supported types of model input and model output. When UE side terminates an AI/ML based/assisted positioning functionality, such as AI/ML assisted positioning with TOA as output, all AI/ML models following this functionality should be deactivated.

For model-ID based LCM of AI/ML based positioning enhancement, an AI/ML model is assigned a unique model ID,which is related to a specific AI/ML model/generalization case/dataset. For example, under an AI/ML based/assisted positioning functionality, an AI/ML model may be developed with a dataset collected in the target area/environment/configuration. When the area/environment/configuration changes, it means that the ongoing AI/ML model may not suitable anymore, and the AI/ML model with the corresponding model ID may be deactivated.

Proposal 47: Functionality based LCM is a coarse-grained type of management, and only static conditions should be included, such as UE capability relate conditions.

Proposal 48: Model-ID based LCM is a fine-grained type of management, and all these dynamic conditions should be included, such as generalization related conditions

Proposal 49: Both functionality based and model-ID based LCM frameworks could coexist and be integrated to acheve flexbile LCM of AI/ML model.

3. Conclusions

In this contribution, we discuss AI/ML based positioning accuracy enhancement with the following observations and proposals.

Proposal 1: Support time domain CIR as one model input for training of AI/ML model for positioning.

Proposal 2: When the entity conducting model training and data collection is not the same, collected data should be delivered from the data-collection entity to the model-training entity.

Proposal 3: For ground truth label collection, to improve the quality of labels, indicate UE the criteria or requirement for data labeling or indicate UE to report label quality indicator.

Proposal 4: For measurement collection, at least measurement type and format should be indicated, depending on the model input.

Proposal 5: In addition to PRU, UE is also used to perform data collection for both label and measurement.

Proposal 6: Further study the specification impact of data collection for semi-supervised learning.

Proposal 7: Real-time on-device model training with a large-scale dataset should be avoided at UE side.

Proposal 8: Other potential issues on data quality and terminal capability should be considered for data collection.

Observation 1: Model transfer over air interface can be achieved with extremely low signaling overhead by combining small-parameter model design and advanced model quantization technologies.

Proposal 9: Further study the overhead of model transfer, and support model transfer over air interface for AI/ML based positioning.

Proposal 10: For the case where model is developed at network side and deployed at UE side, network side should transfer the model information to the target UE.

Proposal 11: Model information should contain meta-information indicating model capability and the physical and network environment or condition under which the model is suitable for operation.

Proposal 12: **The process of model activation and deactivation is needed to flexibly control the model's lifecycle, so as to ensure positioning performance.**

Proposal 13: **Network side should deploy a model pool contraining multiple models with same structure but different parameters to UE in advance for model selection.**

Proposal 14: **Network side could send a model selection instruction to instruct the target UE to select a suitable model from the model pool, when the current model does not work well.**

Proposal 15: **Model monitoring can be achieved based on the following ideas.**

- **a) Monitor covariate (input) shift: detecting whether or to what extent the distribution of model inputs in the test dataset is consistent with the training dataset.**
- **b) Monitor concept (translation) shift: detecting whether or to what extent the mapping between model inputs and model outputs in the test dataset is consistent with the training dataset.**

Observation 2: **Monitoring convariate (input) shift can acheve model input based model monitoring, and can be used as an indirect indicator of degraded performance due to its accessibility without the needs of model output related information, such as ground truth labels.**

Observation 3: **Monitoring concept (translation) shift can acheve model output based model monitoring, and can provide a definite answer on whether the deployed model is still accurate or not, with the needs of model output related information, such as ground truth label.**

Proposal 16: **Multiple monitoring methods can be integrated together to make a model monitoring decision with the considerations of overhead and accuracy.**

Proposal 17: **Study specific model monitoring schemes and their specification impacts for both model input based model monitoring and model output based model monitoring with considering the shift of covariate (input) and concept (translation).**

Proposal 18: **The classification criteria for self-monitoring and non-self-monitoring methods is whether the entire process of model monitoring is implemented on the physical entity where the AI/ML model is deployed without the assistance of other physical entities.**

Proposal 19: **According to the functionality involved in the process of model monitoring, model monitoring can be divided into three entities:**

- **Entity of data collection for model monitoring**
- **Entity of metric calculation for model monitoring**
- **Entity of monitoring decision for model monitoring**

Proposal 20: **For model monitoring, the specific mapping between the logic entities to the physical entities is case-by-case.**

Proposal 21: **Dedicated reference signal may be configured to support data collection for model monitoring.**

Proposal 22: **Under the general model monitoring framework, the further analysis of specific specification impact on model monitoring should be case-by-case, including the mapping between the logic entities to the physical entities and the specific signaling.**

Proposal 23: **Further study specification impact of the shift detection of dominant feature distribution based model monitoring, may include**

- **a) the type(s) of dominant feature(s) for input data**
- **b) the reference distribution(s) of dominant feature(s)**

Proposal 24: **Further study specification impact of AI/ML based adversarial validation based model monitoring, may include**

- **a) The input data collection for monitoring entities for adversarial validation**

Proposal 25: **During training the AI/ML model for positioning, a seperate model could be tailored to continually detect whether the model input is out-of-distribution, so as to monitor the performance of the AI/ML model for positioning, with limited resource consumption.**

Proposal 26: **Further study specification impact of AI/ML based out-of-distribution based model monitoring, may include:**

- **a) AI/ML model for monitoring tranfered from NW side to UE side.**

Proposal 27: **The main specification impact of ground truth label based model monitoring is the procedure of collecting the samples with ground truth labels, which could reuse that of data collection for training.**

Proposal 28: **Further study specification solution and impact on model monitoring for AI/ML assisted positioning, may include**

- **a) The error threshold between AI output and estimated timing based on calculated location**
- **b) The timing measurement enhancement indication based AI/ML assisted positioning**

Proposal 29: **Further study specification impact of motion sensors assisted model monitoring considering the UE privacy, may include:**

- **a) Indication to inform UE side to collect motion sensor information.**

Proposal 30: **During training the AI/ML model for positioning, a seperate ranging model could be tailored to continually monitor the performance of the AI/ML model for positioning, with limited resource consumption.**

Observation 4: **With the assistance of the ranging model, efficient model monitoring can be achieved without the needs of other measurement related to model output.**

Proposal 31: **Further study specification impact of ranging model assisted model monitoring, may include：**

- **a) Ranging model transferred from NW side to UE side for monitoring.**

Proposal 32: **When fine-tuning is conducted at UE side, UE capability corresponding to fine-tuning is required.**

Proposal 33: **To enable model fine-tuning when AI/ML model inference is at UE side, support assistance information to the target UE about pre-trained model and training configuration.**

Proposal 34: **Training data collection request for model fine-tuning and feedback from the target UE is required to support model fine-tuning at network side.**

Proposal 35: **The result of model monitoring and the achievability of model updating should be jointly considered as the condition of model updating.**

Proposal 36: **Support time domain CIR/PDP as model input for AI/ML based positioning.**

Proposal 37: **For direct AI/ML positioning, when model inference is at network side, request to and feedback from the target UE of the necessary measurement (e.g., as the input to the AI/ML model) for model inference is needed.**

Proposal 38: **For AI/ML assisted & UE assisted positioning, support the target UE to report the output of AI/ML model inference (intermediate feature for positioning) when model inference is at UE side.**

Proposal 39: **For AI/ML assisted positioning, when model inference is at network side, request to and feedback from the target UE of the necessary measurement (e.g., as the input to the AI/ML model) for model inference is needed.**

Proposal 40: **TRP-related information, such as encoded TRP ID and TPR's location, should be incorporated into the model input for Construction 2 (Single model, same TRP for N model) of AI/ML assisted positioning without additional signaling overhead.**

Proposal 41: **For Construction 1 (single TRP, N model for N TRPs), uplink positioning should be primarily considered, and AI/ML models should be deployed at gNB side.**

Proposal 42: **For Construction 2 (single TRP, same model for N TRPs), downlink positioning should be primarily considered, and AI/ML model should be deployed at UE side.**

Proposal 43: **For Construction 3 (Multi-TRP, one model for N TRPs), uplink positioning should be primarily considered, and AI/ML model should be deployed at LMF side.**

Proposal 44: **Support TOA as a new measurement reporting from UE side to NW side.**

Proposal 45: **A general model management procedure should be specially studied for AI/ML based positioning accuracy enhancement.**

Proposal 46: **Support to study the detailed assistance signaling configuration when the model management procedure for AI/ML based positioning is agreed.**

Proposal 47: Functionality based LCM is a coarse type of management, and only static conditions should be included, such as UE capability related conditions.

Proposal 48: Model-ID based LCM is a fine type of management, and all these dynamic conditions should be included, such as generalization related conditions

Proposal 49: Both functionality based and model-ID based LCM frameworks could coexist and be integrated to acheve flexbile LCM of AI/ML model.

References

[1] 3GPP RP-213599, "New SI: Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface", #94e, Dec.6-17, 2021.
[2] vivo, R1-2304475, "Evaluation on AI/ML for positioning accuracy enhancement", RAN1 #113, May 22nd – May 26th, 2023.
[3] Nokia, R3-222653, "(TR for TR 37.817) General Framework", #115-e, Februray 21st– March 3rd, 2021.
[4] Quinonero-Candela, Joaquin, et al., eds. Dataset shift in machine learning. Mit Press, 2008.
[5] 3GPP, TS38.305, V17.3.0 , 2022-12.

Appendix A

RAN1 #109e

At the RAN1 #110 meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

Study further on sub use cases and potential specification impact of AI/ML for positioning accuracy enhancement considering various identified collaboration levels.

- Companies are encouraged to identify positioning specific aspects on collaboration levels if any in agenda 9.2.4.2.
- Note1: terminology, notation and common framework of Network-UE collaboration levels are to be discussed in agenda 9.2.1 and expected to be applicable to AI/ML for positioning accuracy enhancement.
- Note2: not every collaboration level may be applicable to an AI/ML approach for a sub use case

Agreement

For further study, at least the following aspects of AI/ML for positioning accuracy enhancement are considered.

- Direct AI/ML positioning: the output of AI/ML model inference is UE location
  - E.g., fingerprinting based on channel observation as the input of AI/ML model
  - FFS the details of channel observation as the input of AI/ML model, e.g. CIR, RSRP and/or other types of channel observation
  - FFS: applicable scenario(s) and AI/ML model generalization aspect(s)
- AI/ML assisted positioning: the output of AI/ML model inference is new measurement and/or enhancement of existing measurement
  - E.g., LOS/NLOS identification, timing and/or angle of measurement, likelihood of measurement
  - FFS the details of input and output for corresponding AI/ML model(s)
  - FFS: applicable scenario(s) and AI/ML model generalization aspect(s)
- Companies are encouraged to clarify all details/aspects of their proposed AI/ML approaches/sub use case(s) of AI/ML for positioning accuracy enhancement

Agreement

Companies are encouraged to study and provide inputs on potential specification impact at least for the following aspects of AI/ML approaches for sub use cases of AI/ML for positioning accuracy enhancement.

- AI/ML model training
  - training data type/size
  - training data source determination (e.g., UE/PRU/TRP)
  - assistance signalling and procedure for training data collection
- AI/ML model indication/configuration

- assistance signalling and procedure (e.g., for model configuration, model activation/deactivation, model recovery/termination, model selection)
- AI/ML model monitoring and update
  - assistance signalling and procedure (e.g., for model performance monitoring, model update/tuning)
- AI/ML model inference input
  - report/feedback of model input for inference (e.g., UE feedback as input for network side model inference)
  - model input acquisition and pre-processing
  - type/definition of model input
- AI/ML model inference output
  - report/feedback of model inference output
  - post-processing of model inference output
- UE capability for AI/ML model(s) (e.g., for model training, model inference and model monitoring)
- Other aspects are not precluded
- Note: not all aspects may apply to an AI/ML approach in a sub use case

Note2: the definitions of common AI/ML model terminologies are to be discussed in agenda 9.2.1

RAN1 #110

At the RAN1 #110 meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

For characterization and performance evaluations of AI/ML based positioning accuracy enhancement, the following two AI/ML based positioning methods are selected.

- Direct AI/ML positioning
- AI/ML assisted positioning
- Note 1: the selection does not intend to provide any indication of the prospects of any future normative project.
- Note 2: further discussion (including selection of other sub use cases and/or down selection of selected sub use cases) are not precluded based on performance evaluation and potential specification impact study results

Agreement

Regarding data collection for AI/ML model training, to study and provide inputs on potential specification impact at least for the following aspects of AI/ML based positioning accuracy enhancement

- Ground truth label determination (e.g., based on UE/PRU/TRP measurement/report)
  - Partial and/or noisy ground truth label
- Signaling for data collection
- Other aspects are not precluded

Agreement

Regarding AI/ML model monitoring and update, to study and provide inputs on potential specification impact at least for the following aspects of AI/ML based positioning accuracy enhancement

- AI/ML model monitoring performance metrics
- Condition of AI/ML model update
- Reference signals and measurement feedback/report
- Other aspects are not precluded

Agreement

Study aspects in terms of potential benefit(s) and requirement(s)/specification impact(s) of AI/ML model training and inference in AI/ML for positioning accuracy enhancement considering at least

- UE-side or Network-side training
- UE-side or Network-side inference
  - Note: model inference at both UE and network side is not precluded where proponent(s) are encouraged to clarify their AI/ML approaches

Note: companies are encouraged to clarify aspects of their proposed AI/ML approaches for positioning

when AI/ML model training and inference are not performed at the same entity

Conclusion

To use the following terminology defined in TS 38.305 when describe their proposed positioning methods

- UE-based
- UE-assisted/LMF-based
- NG-RAN node assisted

Note: companies are required to clarify their positioning method(s) when their approaches do not fall in one of the above

RAN1 #110b

At the RAN1 #110-bis meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

- Study and provide inputs on benefit(s) and potential specification impact at least for the following cases of AI/ML based positioning accuracy enhancement
  - Case 1: UE-based positioning with UE-side model, direct AI/ML or AI/ML assisted positioning
  - Case 2a: UE-assisted/LMF-based positioning with UE-side model, AI/ML assisted positioning
  - Case 2b: UE-assisted/LMF-based positioning with LMF-side model, direct AI/ML positioning
  - Case 3a: NG-RAN node assisted positioning with gNB-side model, AI/ML assisted positioning
  - Case 3b: NG-RAN node assisted positioning with LMF-side model, direct AI/ML positioning

Agreement

Regarding AI/ML model indication[/configuration], to study and provide inputs on potential specification impact at least for the following aspects on conditions/criteria of AI/ML model for AI/ML based positioning accuracy enhancement

- Validity conditions, e.g., applicable area/[zone/]scenario/environment and time interval, etc.
- Model capability, e.g., positioning accuracy quality and model inference latency
- Conditions and requirements, e.g., required assistance signalling and/or reference signals configurations, dataset information
- Note: other aspects are not precluded

Agreement

Regarding AI/ML model monitoring for AI/ML based positioning, to study and provide inputs on potential specification impact for the following aspects

- Assistance signaling and procedure at least for UE-side model
- Report/feedback and procedure at least for Network-side model
- Note1: study is applicable to both of the following cases
  - Model inference and model monitoring at the same entity
  - Entity to perform the model monitoring is not the same entity for model inference
- Note2: other aspects are not precluded

Agreement

Regarding data collection for AI/ML model training for AI/ML based positioning, at least for each of the agreed cases (Case 1 to Case 3b)

- Study whether (and if so how) an entity can be used to obtain ground truth label and/or other training data
  - Companies are requested to report their assumption of the entity (or entities) used to obtain ground truth label and/or other training data for each case (Case 1 to Case 3b)
  - Companies are requested to report their assumption of applicable ground truth label (e.g., location or other information) and/or other training data (e.g., measurement) for each case (Case 1 to Case 3b)

- Feasibility study on the entity to obtain ground truth label and/or other training data takes into account at least
    - availability of the entity to obtain label and/or other training data
- Note: further discussion and decision of the entity (or entities) used to obtain ground truth label and/or other training data for each case (Case 1 to Case 3b) is not precluded based on companies' input

- Study potential signalling and procedure to enable data collection
    - Potential specification impact on the details of request/report of label and/or other training data, and to enable delivering the collected label and/or other training data to the training entity when the training entity is not the same entity to obtain label and/or other training data
    - Potential specification impact on assistance signaling indicating reference signal configuration(s) to derive label and/or other training data

RAN1 #111

At the RAN1 #111 meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

For the study of benefit(s) and potential specification impact for AI/ML based positioning accuracy enhancement, one-sided model whose inference is performed entirely at the UE or at the network is prioritized in Rel-18 SI.

Agreement

Regarding AI/ML model inference, to study and provide inputs on potential specification impact (including necessity and applicability of specifying AI/ML model input and/or output) at least for the following aspects for each of the agreed cases (Case 1 to Case 3b) in AI/ML based positioning accuracy enhancement

- Types of measurement as model inference input
    - new measurement
    - existing measurement
- UE is assumed to perform measurement as model inference input for Case 1, Case 2a and Case 2b; TRP is assumed to perform measurement as model inference input for Case 3a and Case 3b
    - Report of measurements as model inference input to LMF for LMF-side model (Case 2b and Case 3b)
- For AI/ML assisted positioning, new measurement report and/or potential enhancement of existing measurement report as model output to LMF for UE-assisted (Case 2a) and NG-RAN node assisted positioning (Case 3a)
- Assistance signaling and procedure to facilitate model inference for both UE-side and Network-side model
    - New and/or enhancement to existing assistance signaling
    - Note: whether such assistance signaling and procedure can be applied to other aspect(s) of AI/ML model LCM can also be discussed

Agreement

Regarding data collection for AI/ML model training for AI/ML based positioning,

- The following options of entity and mechanisms to generate ground truth label are identified for further study
    - For direct AI/ML positioning, ground truth label is UE location
        - PRU with known location
        - UE generates location based on non-NR and/or NR RAT-dependent positioning methods
        - LMF generates UE location based on positioning methods
        - LMF with known PRU location
        - Note: user data privacy needs to be preserved
    - For AI/ML assisted positioning, ground truth label is one or more of the intermediate parameter(s) corresponding to AI/ML model output
        - PRU generates label directly or calculates based on measurement/location
        - UE generates label directly or calculates based on measurement/location
        - Network entity generates label directly or calculates based on measurement/location

- The following options of entity to generate other training data at least measurement corresponding to model input are identified for further study
    - For UE-based with UE-side model (Case 1) and UE-assisted positioning with UE-side (Case 2a) or LMF-side model (Case 2b)
        - PRU
        - UE
    - For NG-RAN node assisted positioning with Network-side model (Case 3a and Case 3b)
        - TRP
    - Note: other options of entity to generate other training data are not precluded
- Note: Existing PRU definition is in 38.305

Agreement

Regarding data collection for AI/ML model training for AI/ML based positioning, study benefits, feasibility and potential specification impact (including necessity) for the following aspects

- Request/report of training data
    - Ground truth label
    - Measurement corresponding to model input
    - Associated information of ground truth label and/or measurement corresponding to model input
- Assistance signaling and procedure to facilitate generating training data
    - Reference signal (e.g., PRS/SRS) configuration(s) and configuration identifier
    - Assistance information, e.g., between LMF and UE/PRU, for label calculation/generation, and label validity/quality condition, etc.
    - Note1: whether such assistance signaling and procedure can be applied to other aspect(s) of AI/ML model LCM can also be discussed
- Note2: Study may consider different entity to generate training data as well as different types of training data when applicable
- Note3: study considers both of the following cases when applicable
    - when the training entity is the same entity to generate training data
    - when the training entity is not the same entity to generate training data

Agreement

- Regarding AI/ML model monitoring for AI/ML based positioning, to study and provide inputs on feasibility, potential benefits (if any) and potential specification impact at least for the following aspects
    - At least the following are identified for further study as potential data for calculating monitoring metric
        - If monitoring based on model output
            - E.g. , estimated UE location corresponding to model output for direct AI/ML positioning, estimated intermediate parameter(s) corresponding to model output for AI/ML assisted positioning, ground truth label corresponding to model inference output for both direct and AI/ML assisted positioning
        - If monitoring based on model input
            - E.g., measurement corresponding to model inference input
        - Note1: other type of potential data for model monitoring is not precluded
        - Note2: combination of one or more type of potential data for monitoring is not precluded
    - If a given type of data is necessary for calculating monitoring metric, study whether and if so
        - How an entity can be used to provide the given type of data for calculating monitoring metric
            - Companies are requested to report their assumption of the entity (or entities) used to provide the given type of data for calculating monitoring metric for each case
        - Potential signalling for provisioning of the given type of data for calculating associated monitoring metric
        - Potential assistance signaling and procedure to facilitate an entity providing data for calculating monitoring metric
        - Potential UE-network interaction
            - E.g., model monitoring decision indication between UE and network

Agreement

For AI/ML based positioning accuracy enhancement, direct AI/ML positioning and AI/ML assisted positioning are selected as representative sub-use cases.

RAN1 #112

At the RAN1 #112 meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

Regarding training data generation for AI/ML based positioning,

- The following options of entity and mechanisms to generate ground truth label are identified
  - At least PRU is identified to generate ground truth label for UE-based positioning with UE-side model (Case 1) and UE-assisted positioning with UE-side model (Case 2a)
  - At least LMF with known PRU location is identified to generate ground truth label for UE-assisted/LMF-based positioning with LMF-side model (Case 2b) and NG-RAN node assisted positioning with LMF-side model (Case 3b)
  - At least network entity with known PRU location is identified to generate ground truth label for NG-RAN node assisted positioning with gNB-side model (Case 3a)
  - FFS whether and if so, applicable conditions and potential specification impact for the following options to generate ground truth label
    - UE generates ground truth label based on non-NR and/or NR RAT-dependent positioning methods
    - Network entity generates ground truth label based on positioning methods
- The following options of entity to generate other training data (at least measurement corresponding to model input) are identified
  - For UE-based with UE-side model (Case 1) and UE-assisted positioning with UE-side (Case 2a) or LMF-side model (Case 2b)
    - PRU
    - UE
  - For NG-RAN node assisted positioning with Network-side model (Case 3a and Case 3b)
    - TRP
- Note: transfer of training data from the entity generating training data to a different entity is not precluded and associated potential specification impact is for further study

Agreement

Regarding training data collection for AI/ML based positioning, study benefit(s) and potential specification impact (including necessity) at least for the following aspects

- Associated information of training data
  - Quality indicator at least for ground truth label (if needed)
  - Other information associated with training data is not precluded. E.g., information related training dataset/samples, information related to scenario, resource configuration & mapping, timing for training data, information on implementation imperfections, etc.
- Assistance signaling and procedure to facilitate generating/collecting training data
  - Potential determination of the UE/PRU/TRP which can provide the training data
  - Configuration of reference signal (for measurement and/or label)
  - Signaling other than above 2 for data collection
    - E.g., requested quality of training data

Agreement

Regarding AI/ML model monitoring for AI/ML based positioning, to study and provide inputs on benefit(s), feasibility, necessity and potential specification impact for the following aspects

- Entity to derive monitoring metric
  - UE at least for Case 1 and 2a (with UE-side model)
    - FFS PRU for Case 1 and 2a
  - gNB at least for Case 3a (with gNB-side model)
    - FFS gNB for Case 3b (with LMF-side model)
  - LMF at least for Case 2b and 3b (with LMF-side model)
  - Note1: companies are requested to report their assumption of entity to calculate monitoring metric if different from above options for each of the agreed cases (Case 1 to Case 3b)
- If model monitoring does not require ground truth label (or its approximation).
  - Monitoring metric, e.g., statistics of measurement, relative displacement, inference output inconsistency, etc.

    - Assistance signaling and procedure, e.g., RS configuration(s) for measurement, measurement statistics as compared to the model input statistics of the training data, etc.
    - report of the calculated metric and/or model monitoring decision
- If model monitoring requires and is provided ground truth label (or its approximation)
    - Monitoring metric, e.g., statistics of the difference between model output and ground truth label, etc.
    - Assistance signaling and procedure, e.g., from LMF to UE/gNB indicating ground truth label and/or measurement, etc.
    - report of the calculated metric and/or model monitoring decision
- Note2: other options (of monitoring methods, monitoring metrics, assistance signaling) are not precluded

Agreement

Regarding AI/ML model inference, to study the potential specification impact (including the feasibility, and the necessity of specifying AI/ML model input and/or output) at least for the following aspects for AI/ML based positioning accuracy enhancement

- For direct AI/ML positioning (Case 2b and 3b), type of measurement(s) as model inference input considering performance impact and associated signaling overhead
    - Potential new measurement: CIR/PDP
    - existing measurement: e.g., RSRP/RSRPP/RSTD
    - Note1: details of potential new measurement and/or potential enhancement to existing measurement is to be studied
    - Note2: study the impact of model input for other cases are not precluded
- For AI/ML assisted positioning with UE-assisted (Case 2a) and NG-RAN node assisted positioning (Case 3a), measurement report to carry model output to LMF
    - new measurement report: e.g., ToA, path phase
    - existing measurement report: e.g., RSTD, LOS/NLOS indicator, RSRPP
    - enhancement of existing measurement report: e.g., soft information/high resolution of RSTD
- Assistance signaling and procedure to facilitate model inference for both UE-side and Network-side model
    - RS configurations
    - Other assistance information is not precluded

Note: Companies are encouraged to report their assumption of functionality and their assumption of information element(s) of AI/ML functionality identification for AI/ML based positioning with UE-side model (Case 1 and 2a).

RAN1 #112bis-e

At the RAN1 #112bie e-meeting, some agreements on sub use cases and potential specification impact have been reached, which are listed as follows:

Agreement

Regarding monitoring for AI/ML based positioning, at least the following entities are identified to derive monitoring metric

- UE at least for Case 1 and 2a (with UE-side model)
- gNB at least for Case 3a (with gNB-side model)
- LMF at least for Case 2b and 3b (with LMF-side model)

Working Assumption

Regarding data collection at least for model training for AI/ML based positioning, at least the following information of data with potential specification impact are identified.

- Ground truth label
    - At least for model training

    - Report from the label data generation entity
- Measurement (corresponding to model input)
    - At least for model training
    - Report from the measurement data generation entity
- Quality indicator
    - For and/or associated with ground truth label and/or measurement at least for model training
    - Report from the label and/or the measurement data generation entity and/or as request from a different (e.g., data collection, etc.) entity
- RS configuration(s)
    - At least for deriving measurement
    - Request from data generation entity (UE/PRU/TRP) to LMF and/or as LMF assistance signaling to UE/PRU/TRP
    - Note1: there may not be any enhancements on top of existing RS configuration(s) or any new RS configuration(s) for positioning measurement
- Time stamp
    - At least for and/or associated with training data for model training
        - Separate time stamp for measurement and ground truth label, when measurement and ground truth label are generated by different entities
    - Report from data generation entity together with training data and/or as LMF assistance signaling
    - Note2: there may not be any enhancements on top of time stamp in existing positioning measurement report or any new time stamp report for positioning measurement
- FFS other necessary information (e.g., scenario identifier. LOS/NLOS condition, timing error, etc.) for data collection
- Note3: whether the above information can be applied to other aspects of AI/ML LCM (e.g., updating, monitoring, etc.) can also be discussed
- Note4: transfer of data from the entity generating data to a different entity is not precluded from RAN1 perspective

Agreement

Regarding monitoring for AI/ML based positioning, at least the following aspects are identified for further study on benefit(s), feasibility, necessity and potential specification impact for each case (Case 1 to 3b)

- Assistance signaling from LMF to UE/PRU/gNB for UE/gNB-side model monitoring
- Assistance signaling from UE/PRU for network-side model monitoring
- Model monitoring based on provided ground truth label (or its approximation)
    - Monitoring metric: statistics of the difference between model output and provided ground truth label
    - Provisioning of ground truth label and associated label quality
- Model monitoring using at least statistics of measurement(s) without ground truth label
    - Monitoring metric: e.g., statistics of measurement(s) compared to the statistics associated with the training data
    - Note1: the measurement(s) may or may not be the same as model input
- Note2: other monitoring methods (e.g., based on statistics of model output without ground truth label, based UE motion sensor and/or jointly based on multiple monitoring metrics) are not precluded

Agreement

Regarding LCM of AI/ML based positioning accuracy enhancement, at least for Case 1 and Case 2a (model is at UE-side), further study the following aspects on information related to the conditions

- What are the conditions for functionality-based LCM
    - which aspects should be specified as conditions of a Feature/FG available for functionality
- What are the conditions for model-ID-based LCM
    - Which aspects should be considered as additional conditions, and how to include them into model description information during model identification