|  |  |
| --- | --- |
|   | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

|  |  |
| --- | --- |
|  | **Public Document** |
| **N611** | 22/03/22 |
| **Source** | Requirements (NNW) |
| **Title** | MPAI-NNW Use cases and functional requirements – draft 1.0  |
| **Target** | MPAI Community |

# Introduction

During the last decade, Neural Networks are deployed in an increasing variety of domains, but solution and especially deep neural network are costly. The process of AI training is costly not only in terms of resources (GPUs, CPUs, memory) but also time. According to ThinkML, the development of a custom AI solution ranges from $ 6, 000 to $ 300, 000, while renting a pre-built module would cost around $ 40, 000/year. Consequently, it becomes important to guaranty the traceability (owner) and integrity (user) of Neural Networks. Inherited from the multimedia realm, watermarking regroups a family of methodological and applicative tools allowing to **imperceptibly** and **persistently** insert some **metadata** into an original content.

# Purpose of the standard

The current planes for the MPAI Neural Network Watermarking (NNW) standard is to provide the means to measure, for a given size of the watermarking payload applied to the weights of an NN model:

1. The task-dependent impact of the watermark on the performance of the Neural Network.
2. The performance of the watermark detector applied to a modified (e.g., transfer learning, pruning) watermarked network (the performance of the watermark detector is its ability to accurately detect the inserted watermark).
3. The processing cost of watermark injection, e.g., time of execution on a given processing environment, computational complexity.

# Users of watermarking technology for NN

Four types of actors are identified as playing a role in the use cases.

* *NN owner* – the developer of the NN, wishing to ensure that ownership of NN can be claimed.
* *NN watermarking provider* – the developer of the watermarking technology able to carry a payload in a neural network or in an inference.
* *NN customer* – the user who needs the NN owner’s NN to make a product or offer a service.
* *NN end-user* – the user who buys an NN-based product or subscribes to an NN-based service.

# Use cases

The use cases are structured into two categories: the first relates to the NN *per se* (*i.e.*, to the data representation of the model, as discussed in Section 3.1) while the second to the inference (*i.e.*, to the result produced by the network when fed with some input data, as discussed in Section 3.2).

The use cases are presented as sequence diagrams describing the positions and actions of the four main actors in the workflow.

## Use cases related to watermarking the Neural Network model

Two types of use case belong to this category:

* *payload*, *i.e.*, data carried by the watermark is used to identify the actors or the model; this case is presented in Section 3.1.1.
* *loss of integrity*, *i.e.*, data carried by the watermark is used to identify modifications in the model; this case is presented in Section 3.1.2.

### Payload (channel of information)

Data is carried by the watermark is used to identify:

* the ownership of an NN.
* an NN (as if it were a DOI).

#### Identify the ownership of an NN



*Figure 1:* *Identify the ownership of an NN use case: NN owner and NN customer identifiers are inserted*

*Description of Figure 1 workflow:*

* *NN customer* gets needs from product/service from *NN end-users*.
* *NN customer* requests NN model from *NN owner* in order to be able to create the product/service requested by the end user.
* *NN customer* and *NN owner* share the need to protect NN intellectual property; *NN customer* does not want others to use the model to make similar products or offer similar services; *NN owner* wants to acquire others customers as *NN customer*; ideally, the *NN end-user* ID should also be added to the watermark (*cf*. the workflow in Figure 2).
* *NN end-user* acquires the product and/or access to the service with the embedded NN watermark.



*Figure 2:* *Identify the ownership of an NN use case*: in addition to *NN owner* and *NN customer* identifiers, the *NN end-user* identifier is also inserted

*Description of Figure 2 workflow:*

* *NN customer* gets needs from product/service from end-users.
* *NN customer* requests NN model from *NN owner* in order to be able to create the product/service requested by the *NN end-user*.
* *NN customer* and *NN owner* share the need to protect NN intellectual property; *NN customer* does not want other to use the model to make similar products or offer similar services ; *NN owner* wants to acquire other customers as *NN customer*.
* *NN customer* needs to make sure that *NN end-users* do not share the AI solution, thus they insert an identifier for each *NN end-user*.
* *NN end-user* acquires the product and/or access to the service with the embedded NN watermark.

#### Identify an NN (e.g. DOI)



*Figure 3:* *Identify an NN use case*: *NN receives an ID (e.g. DOI)*

*Description of Figure 3 workflow:*

* *NN owner* wants its NN to receive a specific identifier.
* *NN watermark provider* gives a solution with a specific identifier for any new Neural Network and manages the ID usage through its lifecycle (e.g., validation to third parties, or ID record deletion when no longer used).

### Loss of integrity



*Figure 4: Check the NN integrity use case*

*Description of Figure 4 workflow:*

* *NN owner* wants a watermark that permits them to check the integrity of the NN.
* *NN watermark provider* inserts an integrity validation watermark in the NN.
* *NN owner* can distribute the Watermaked NN to their customers.
* *NN owner* can check the integrity and detect modifications of their NN.

## Inference



*Figure 5: Watermarked NN inference use case*

*Description of Figure 5 workflow:*

* *NN customer* gets needs from product/service from end-users.
* *NN customer* requests NN model from *NN owner* in order to be able to create the product/service requested by the end user.
* *NN customer* and *NN owner* share the need to protect NN intellectual property; *NN customer* does not want others to use the model to make similar products or offer similar services; *NN owner* wants to acquire others customers as *NN customer*.
* *NN end-user* can feed the Watermarked NN with input data and receive the inference which is watermarked. The contained ID related information can be the same as in Section 3.1 Use cases related to watermarking the Neural Network model, for instance.

# Conclusion

The use cases identified so far and presented in this document can be structured according to Figure 6.



*Figure 6: tree-representation of the use cases*