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

|  |  |
| --- | --- |
| **N569** | 22/02/23 |
| **Source** | Requirement (NNW) |
| **Title** | The MPAI Neural Network Watermarking (NNW) project |
| **Target** | MPAI Community |

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an international non-profit organisation with the mission to promote the efficient use of data. MPAI develops technical specifications of moving picture, audio and data coding, decoding and technologies that help integration of coding components. Moreover, MPAI aims to bridge the gap between Technical Specifications and their practical use through the development of IPR Guidelines.

In the last decade, Neural Networks have been deployed in an increasing variety of domains. However, the process of AI training is costly not only in terms of resources (GPUs, CPUs, memory) but also time. According to ThinkML [1], the development of a custom AI solution ranges from $ 6.000 to $300.000, while renting a pre-built module may cost around $ 40.000/year. Consequently, for the owner it becomes important to ensure traceability (owner) and for the user integrity of Neural Networks.

Watermarking is a technology inherited from the multimedia domain, regrouping a family of methodological and applicative tools allowing to **imperceptibly** and **persistently** insert some **metadata** into an original content. When applied to the AI world, the following is required:

* Watermarking shall preserve the commercial value/usage of the original network and not affect the performance of the AI usage.
* Metadata shall be recoverable even when the watermarked content was subjected to modifications.
* The inserted metadata generally convey owner and network identity, and usage conditions.

There is significant literature around neural network watermarking presenting different objectives and testing methods. However, any technique can be either **white-box**, when the watermark is embedded inside the parameters of the network or **black-box,** when the watermark is embedded in the inference.

The MPAI Neural Network Watermarking (NNW) project is developing requirements for a future MPAI standard enabling the measure, for a given size of the watermarking payload, of

1. The impact on the performance of the Neural Network.
2. The resistance to modifications, e.g., transfer learning, pruning.
3. The processing cost of watermark injection, e.g., time, processing cost.

This is a new MPAI project and, based on the MPAI process, participation is open to non-members. If you are interested, please contact the MPAI secretariat.

[1] <https://thinkml.ai/is-ai-cost-effective/>