Meta-learning in Deep Neural Networks

One of the main human skills achieved in educational cycles is the ability to learn. First, in school, we learn how to acknowledge knowledge effectively, and then during studies, we use this skill to quickly grasp novel, much more complicated ideas.

The situation is quite different in the case of artificial intelligence (AI), particularly in deep neural networks. While AI performs comparably to humans or surpasses them in many tasks, its ability to learn new tasks based on previously accumulated knowledge is limited and brings many challenges. One of them is catastrophic forgetting, i.e. the tendency for knowledge of the previously learned tasks to be abruptly lost as information relevant to the current tasks. For example, a model trained to recognize 100 bird species based on photos and finetuned for another 50 species will obtain poor performance for the initial 100 species. Another challenge of AI is the need for an enormous training set, while at the same time, humans require only a few examples because they competently use previously accumulated knowledge. The discipline considering those challenges, *meta-learning*, is the main topic of this grant.

In this grant, we would first like to concentrate on introducing novel methods for continual learning and effective training with a limited training set. Then, we plan to provide explainable and interpretable technics of meta-learning, which will explain their predictions.

For explainable and interpretable techniques, we plan first to introduce explainability methods that explain the predictions of existing meta-learning approaches, and then we will provide novel self-explainable techniques with a built-in interpretability mechanism. Understanding and controlling AI methods is essential, especially in such crucial applications as medical diagnosis. For instance, medical doctors will obtain an automatic diagnosis and its cause during diagnosis. This way, he will be able to check if the cause of prediction is correct, i.e. the presence of cancer cells, or incorrect, i.e. dust on the camera lens. This way, they will be able to neutralize incorrect automatic diagnosis.

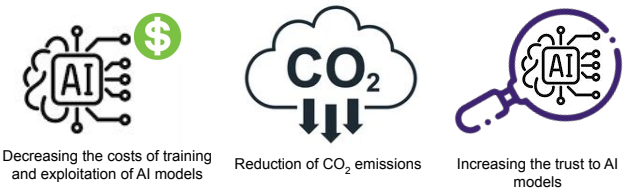

Figure 1: This grant aims to develop meta-learning of deep neural networks in order to decrease the cost of training and exploitation, reduce the carbon dioxide in the atmosphere, and increase the transparency of artificial intelligence.

Concluding, this project aims to achieve breakthrough results in meta-learning, a subfield of machine learning. Its effect will positively impact not only the economy and transparency of deep neural networks but also the natural environment by decreasing the demand for computing power in deep model training.