**Vertical optimization methods for Hardware dimensioning and AI algorithms configuration**

The main purpose of the research project consists of studying and developing methods to automatize the process of matching the configuration of Artificial Intelligence (AI) algorithms with and available hardware resources under user-defined constraints and to solve different tasks from various domain. This topic is in line with the goal of the StairwAI EU Horizon2020 project, I.e., to develop an optimization engine for matching target AI applications with the HW platforms (from edge to cloud and on a wide range of architectures).

Generally speaking, matching available computing resources and algorithm configurations under solution time and quality constraints is a complex task, especially when multiple algorithms can be used for solving the same problem. Often, a key requirement is recommending users the proper technique for the application (e.g., suggesting the right algorithm) and also finding the proper hardware (HW) resources to run it, together with the optimal configuration.

The project will consist of two main phases. At first, multiple Machine Learning (ML) models will be trained to learn the relationship between AI algorithms (under various configurations) and their performance (e.g., runtime, memory footprint, solution quality, etc.). The models will be trained using a combination of already existing data sets from to the energy system domain and new data sets to be created during the project (by benchmarking different AI algorithms on a heterogeneous set of HW architectures). In the second phase of the project, the ML algorithm will be integrated with an optimization model to automatically recommend a portfolio of solutions (algorithms and their configurations) given user constraints and available HW (and budget).

**Studio di metodi per ottimizzazione verticale applicata al dimensionamento di risorse hardware e configurazione di algoritmi di Intelligenza Artificiale**

Lo scopo principale del progetto di ricerca consiste nello studio e nello sviluppo di metodi per automatizzare il processo di corrispondenza della configurazione di algoritmi di Intelligenza Artificiale (AI) con e risorse hardware disponibili sotto vincoli definiti dall'utente e per risolvere diversi compiti da vari domini. Questo argomento è in linea con l'obiettivo del progetto StairwAI EU Horizon2020, ovvero sviluppare un motore di ottimizzazione per abbinare le applicazioni AI target con le piattaforme HW (dall'edge al cloud e su un'ampia gamma di architetture).

In generale, abbinare le risorse di calcolo disponibili e le configurazioni degli algoritmi in condizioni di tempo di soluzione e vincoli di qualità è un compito complesso, soprattutto quando è possibile utilizzare più algoritmi per risolvere lo stesso problema. Spesso, un requisito fondamentale è consigliare agli utenti la tecnica corretta per l'applicazione (ad esempio, suggerire l'algoritmo giusto) e anche trovare le risorse hardware (HW) appropriate per eseguirla, insieme alla configurazione ottimale.

Il progetto si articolerà in due fasi principali. Inizialmente, verranno addestrati più modelli di Machine Learning (ML) per apprendere la relazione tra gli algoritmi di intelligenza artificiale (in varie configurazioni) e le loro prestazioni (ad es. tempo di soluzione, consumo di memoria, qualità della soluzione, ecc.). I modelli saranno addestrati utilizzando una combinazione di set di dati già esistenti dal dominio del sistema energetico e nuovi set di dati da creare durante il progetto (tramite il benchmarking di diversi algoritmi di intelligenza artificiale su un insieme eterogeneo di architetture HW). Nella seconda fase del progetto, l'algoritmo ML sarà integrato con un modello di ottimizzazione per consigliare automaticamente un portafoglio di soluzioni (algoritmi e loro configurazioni) dati i vincoli dell'utente e l'HW (e budget) disponibile.