Executive Summary of the Execution Configuration Guide

Funded by the European Union - NextGenerationEU. However, the views and opinions expressed are solely those of the author(s) and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

The *Execution Configuration Guide* provides a practical methodological framework for **configuring, executing, and operating artificial intelligence models in an efficient, reproducible, and sustainable manner**, with a strong focus on reducing energy consumption and complying with European regulatory requirements. The guide is aimed at developers, engineering teams, infrastructure managers, and AI project leaders seeking to implement best practices across the full lifecycle, from development to deployment and operation.

The document defines a set of **core principles** for AI execution, including energy transparency, reproducibility, dependency traceability, and proactive model optimization prior to deployment. It highlights the importance of documenting energy consumption in line with emerging European regulations and ensuring consistent execution environments through containerization, version control, and automation using DevOps and CI/CD practices.

A key pillar of the guide is the **joint optimization of software and hardware**, covering infrastructure selection (cloud versus on-premise), efficient use of accelerators such as GPUs and TPUs, optimized libraries, and dynamic performance tuning techniques. The guide also examines different **distributed execution approaches**, including data parallelism, model parallelism, and federated learning, providing clear criteria to assess when these architectures are beneficial and when they may introduce unnecessary energy overhead.

Finally, the guide includes **practical checklists, actionable recommendations, and a real-world application example**, supporting the effective adoption of these practices in production environments. Overall, the document serves as a reference to ensure that AI model execution is technically robust, energy-efficient, and aligned with sustainability, governance, and responsible digital transformation principles.