CERN IT Machine Learning Infrastructure Workshop - Summary

March 10 2023

Sofia Vallecorsa, Ricardo Rocha

Summary

5 presentations from different IT services / areas

7 short presentations from different ML applications in CERN IT

Final discussion on common areas of work and actions

Actions

→**Propose ML initiatives with all business engagement units**

    To help track requirements from the different communities

    To assert the offering in CERN IT is aligned with those requirements

→**Improve GPU sharing between different IT services**

    Given the scarcity of this type of resource

    With the short-term goal of optimizing the overall usage of the existing infrastructure

    Note: A short term task, split from the goal of thinking the future ML infrastructure

Actions

→**Investigate integration of public cloud resources**

    Low level (GPUs) as well as higher level services

    Consider usage reporting, accounting, etc

→**Ensure ML needs are taken into account in long term resource procurement**

    Access to public cloud services, procurement of accelerators

→**Improve dataset and model registry and bookkeeping**

    Both local and external (reusable) models

Actions

→**Organize a second workshop**

    Focus on use case and requirements CERN wide

    Build on feedback from the different business engagement initiatives, and others