

Target X Evaluating infrastructure and platforms for AI/ML R&D

Duration: 12 months

Description: We can anticipate a rapid expansion of research projects in the Nordics in the general area of artificial intelligence, in particular in machine learning and deep learning. Frameworks such as Tensorflow and high-level libraries such as Keras have become massively popular, with applications across many areas. The work will be divided into two tracks, resulting in a concerted effort in transforming the experience gained into reusable material for installing, production setting and porting AI workflows into different Nordic infrastructures.

Track1 Domain specific workflows will be enabled on the NIRD Toolkit, to support the use-cases described below. Tests will be done to verify the portability of the workflows from one Nordic infrastructure to another, thus paving the way towards a possible future design of a truly domain specific, platform agnostic AI/ML infrastructure.

Track2 In this track we will evaluate whether the rapidly expanding open source project KubeFlow which can be an efficient means of providing both ML and DevOps infrastructure for AI research projects such as HASTE <http://haste.research.it.uu.se/>, funded by the Swedish foundation for strategic research (SSF). The goal is to develop a smart system for handling large amounts of image data and central to the project is the notion of proactive storage prioritization driven by machine learning models predicting the relevance of new data blobs/images.

Track 1 use case(s): We will work at enabling use scenarios from two different domains. The first case is aiming at developing a novel computational platform for designing next-generation therapeutics (NGT), diagnostics and vaccines by combining high-throughput single-cell antibody and T-cell receptor sequencing with artificial intelligence. The second is in the field of neural science and will aim at identifying infected cell from deep learning analysis on data from measurement of spiking activities in neural cells.

Track 2 use case: The HASTE software needs an infrastructure to train, test and deploy ML models in production. We will apply the developed KubeFlow environments to the HASTE development process (<https://github.com/HASTE-project>). CSC evaluates the KubeFlow environments with a use case of automated workflows for annotation of digitised archive materials.

Deliverables:

D1: Install and optimize the DL platform for enabling the neural diagnostic workflows.

D2: Install and optimize the AI/ML platform for enabling NGT AI/ML workflows.

D3: A working KubeFlow environment deployed and demonstrated over a virtual Kubernetes cluster in at least one of SNIC Science Cloud or cPouta.

D4: A brief report summarizing the experiences of working with Kubeflow, illustrated through the HASTE case study, and including recommendations for future work.