AI Lab – An Ecosystem for Managing AI projects[1]

Martin Leucker[2], Maria Ostanina[3]

[1] This work was funded by the Bundesministerium für Bildung und Forschung (BMBF) through the KI-Lab Lübeck Project.

[2] Universität zu Luebeck, Institut für Softwaretechnik und Programmiersprachen, Ratzeburger Allee 160, 23562 Luebeck, Deutschland, leucker@isp.uni-luebeck.de

[3] Universität zu Luebeck, Institut für Softwaretechnik und Programmiersprachen, Ratzeburger Allee 160, 23562 Luebeck, Deutschland, maria.ostanina@isp.uni-luebeck.de

Abstract: This paper presents AI Lab, an ecosystem with tools and resources for building AI-based systems. It provides a basic infrastructure for building and implementing AI systems rapidly, with a focus on machine learning capabilities (challenges/ issues), i.e. training and running models, storing, analyzing, and processing data. More over, it supports the deployment of the built solutions on different target systems including a broad range of edge systems. The AI Lab is designed to serve both novices as well as experts in AI by knowledge transfer from university to practice, spanning from bachelors to researchers.

Keywords: artificial intelligence platform; integrated development platform

1 Introduction

The AI Lab Lübeck forms the basis of an ecosystem supporting the rapid development of machine learning-based solutions. While it supports educational purposes, it is especially suitable for research projects, and, in the long run, also for professional projects. It is built up by high-performance machine learning computers for high-end learning tasks as well as midsize servers for interactive development. Regarding software, it integrates the most popular open source software solutions for machine learning. Moreover, it offers a configurable deployment tool chain allowing to employ the learning results on the variety of edge devices. For smooth management of machine learning based projects it offers various collaboration capabilities, like cloud data storage and basic project management tools, an open data portal etc. The lab is accessible via ai-lab.digital-hub-luebeck.de. Figure 1 shows screen-shot of the main page of the lab.

The anticipated or realized functionality and design of AI Lab builds upon user studies performed with various project leaders of AI projects at the University of Luebeck and a study of similar labs like Google Colaboratory[1], Peltarion[2], Orange[3], Kaggle[4] respectively.

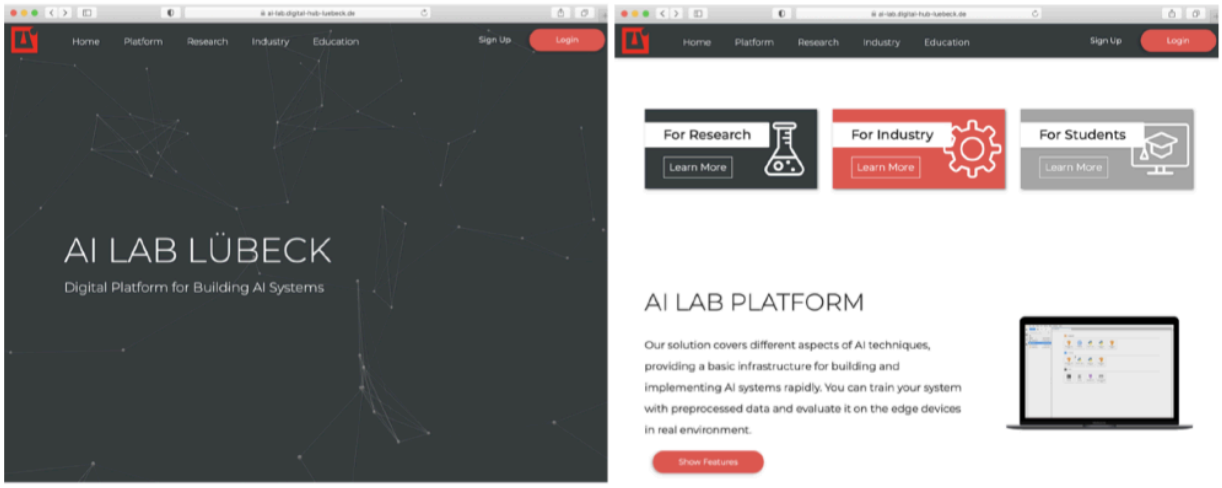

Fig. 1: Screenshots of the AI Lab

The AI Lab ecosystem grows within *KI-Lab Lübeck* project funded with almost 2 Million Euros by the German Ministry of Education and Research (BMBF).

2 The AI Lab from a Users Perspective

The AI Lab is designed to support users within machine learning projects. Such projects typically pursue the following workflow:

1. Data is collected and annotated, checked for quality and possibly enriched, to serve as the basis for machine learning.
2. Using specialized forms self probabilistic differential programming, suitable neural networks are trained and later tested.

[1] https://colab.research.google.com/notebooks/intro.ipynb

[2] https://peltarion.com/

[3] https://orangedatamining.com/

[4] https://colab.research.google.com/notebooks/intro.ipynb

3. If the trade model is considered to be adequate, it is deployed on one or more target platforms.

Currently, mostly image analysis as well as control solutions for robots are developed within the AI Lab.

Let us consider the anticipated workflow in more detail: The user either has or requires a login for the AI Lab platform. This login serves as single-sign-on and allows transparent access to each individual service according to the authorization provided by the AI Lab administrators.

The next step is typically to define a new project and to assign coworkers to the project, if needed. Each project leader can authorize project members to work on the project by defining access rights in a fine granular manner.

A project is composed of storage for data as well as algorithms, the latter, in our lab, typically python programs developed via Jupyter notebooks.

After uploading the corresponding data using the underlying Nextcloud facilities, it is analyzed using several data validation services available in the lab. The AI Lab comes with an open data platform which allows to use public data for the own project but also to share own data under several open data licenses.

Once the data is in reasonable shape,the user creates a Jupyter notebook and specifies the machine learning task. The user typically experiments with the program first on smaller machines but may schedule the program for batch processing on the high-performance computers, once the parameters for the learning program seem to be adequate. Nevertheless, the AI Lab supports persistent sessions allowing to suspend the work also within the interactive usage.

Figure 2 shows two screen-shots of using Juypiter notebook within the lab.

Fig. 2: Screenshots showing the use of the AI Lab

Once a model is successfully trained it may be tested regarding the quality and deployed to target systems. The lab provides the execution of Kubernetes Pods for using the trained model in the cloud but also transformation tools for running the solution on dedicated hardware. Via ONNX we support a variety of different edge platforms, ranging from Nvidia Jetson and similar Intel based solutions to dedicated FPGA boards (especially those from Xilinx).

Whenever several iterations of a typical machine learning workflow will be carried out, for example when continuously enriching the models based on new data, it is helpful to define individual learning workflows. Within the lab this is supported by kubeflow and elyra which provide a graphical user interface for defining complex machine learning workflows.

3 The AI Lab from a Technical Perspective

The AI Lab supports both the storage and the processing of data.

3.1 Storing of data

For storing of data, both a cloud space as well as an open data platform is provided. For the cloud space, AI Lab relies on Nextcloud, an open source cloud solution [Ne21]. Storage devices can be easily integrated into the Nextcloud via the webdav protocol, allowing to scale the data space continuously.

Moreover, we provide an open data portal which relies on the open source framework DKAN [DK21]. Based on the CMS Drupal, it provides both an API as well as a front-end to support batch processing and interactive usage, respectively.

Note that both, the Nextcloud as well as the open data portal (which also store data based on the webdav protocol), is limited in terms of latency and throughput. That means that for machine learning purposes, relevant subsets of the data to be considered should be copied to high-performance storage devices directly attached to machine learning devices (like the ones included in the NVIDIA DGX systems).

3.2 Processing of data

The AI Lab is build with scalability and resilience in mind. For processing of data, the AI Lab provides a variety of different hardware and software solutions. For the hardware part, high-performance systems from NVIDIA (one DGX-2 and one DGX-A100 to be extended by two further DGX-A100) for intensive machine learning jobs extend a set of smaller servers (DELL with GPUs) mainly used for solution engineering and small scale applications. All machines run jupyter notebooks [Ju21] to simplify interactive development. Moreover, we use kubeflow [Ku21] and elyra [El21]to support complex machine learning pipelines, persistent sessions and scheduling.

As software, a set of open source machine learning frameworks such as pytorch [Py21] and tensorflow [Te21] are provided in the lab. The rapid development of new machine learning solutions will be reflected by a constant update and extension of the software frameworks available in AI Lab.

Moreover, the AI Lab supports the development for edge devices. It offers deployment tool chains implementing the ONNX standard [ON21] that enables AI developers to use models with a variety of frameworks, tools, runtimes and compilers.

The overall platform is enriched with monitoring and statistics capabilities to ensure the quality of service of the overall system.

4 Conclusion and Future Work

In this paper, we briefly described the AI Lab currently built up in Lübeck. It supports the rapid and continuous development of machine learning based systems. In the future, we plan to enrich both the hardware of our system but especially also the engineering process of such systems by further methods and methodologies. The AI legislation [EU21] currently under discussion in the EU will require new engineering process support especially for high risk AI-based systems and it is our goal to provide such support in the AI lab in the new future.

References

[DK21] DKAN Community: DKAN Documentation, https://demo.getdkan.org/modules/contrib/dkan/docs/index.html modules/contrib/dkan/docs/index.html, Accessed 2021-05-15, 2021.

[El21] Elyra Team: Elyra Documentation, https://elyra.readthedocs.io/en/latest/getting_started/overview.html, Accessed 2021-05-15, 2021.

[EU21] EU Commission: Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence, https: //ec.europa.eu/commission/presscorner/detail/en/IP_21_1682, Accessed 2021-05-15, 2021.

[Ju21] Jupiter Team: Jupiter Documentation, https://jupyter.readthedocs.io/en/ latest/, Accessed 2021-05-15, 2021.

[Ku21] Kubeflow Authors: Kubeflow Documentation, https://www.kubeflow.org/ docs/, Accessed 2021-05-15, 2021.

[Ne21] Nextcloud: Nextcloud Solution Architecture: Bring data back under control of IT, http://nextcloud.com, Accessed 2021-05-15, 2021.

[ON21] ONNX Team: ONNX Documentation, https://onnx.ai/, Accessed 2021-05-15, 2021.

[Py21] PyTorch Team: PyTorch Documentation, https://pytorch.org/docs/stable/index.html, Accessed 2021-05-15, 2021.

[Te21] TensorFlow Team: TensorFlow Documentation, https://www.tensorflow.org/guide, Accessed 2021-05-15, 2021.

Author information

Prof. Dr. Martin Leucker

Martin Leucker is currently a professor at the University of Lübeck, Germany heading the Institute of Software Engineering and Programming Languages. He obtained his Ph. D. at the RWTH Aachen, Germany and afterwards, he worked as a Postdoc at the University of Philadelphia, USA and at the Uppsala University, Sweden. He pursued his habilitation at the TU München, Germany. He is the author of more than 100 peer reviewed conference and journal papers ranging over software engineering, formal methods and theoretical computer science.

Maria Ostanina

Maria Ostanina is a member of the scientific staff at the Institute for Software Engineering and Programming Languages at the University of Lübeck. She is the project manager of the "KI-LAB Lübeck" project, focusing on building an AI infrastructure and co-developing new educational programs and qualification activities. Her research activities are concerned with building and managing integrative collaborative platforms.