S06 – Develop an MLOps stack

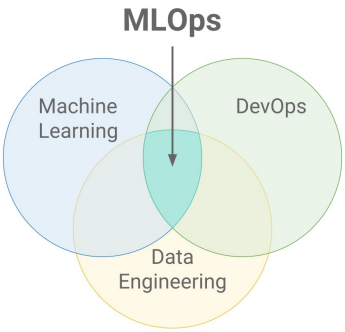
6-month internship @ CEA List

Internship context

Based in Saclay (Essonne), the LIST is one of the two institutes of CEA Tech, the technological research division of the CEA. Dedicated to intelligent digital systems, its mission is to carry out technological developments of excellence on behalf of industrial partners in order to create value.

Within the LIST, the Laboratory of Vision and Learning for Scene Analysis (LVA) conducts research in the field of computer vision and artificial intelligence for the perception of intelligent and autonomous systems. The laboratory's research themes include visual recognition, behavior and activity analysis, large-scale automatic annotation, and perception and decision models. These technologies are applied in major sectors such as security, mobility, advanced manufacturing, healthcare, and sports...

Missions



Source : wikipédia

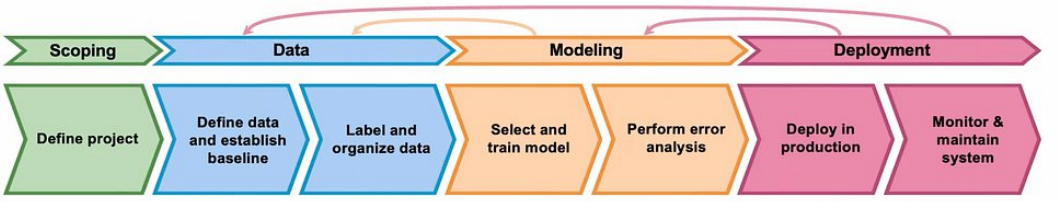
MLOps is a set of good practices to deploy and maintain Machine Learning models in production reliably and efficiently. It is at the frontier between Machine Learning to build models, DevOps to deploy the model and manage the infrastructure, and Data Engineering to handle a large amount of data.

It is a vital part of an healthy development environment of Machine Learning to ensure that the models meet the efficiency and the reliability that they were designed for.

In order to develop scene and text comprehension tools, the SIALV (Vision and Language Artificial Intelligence Service) in which the LVA inhabit is creating day to day algorithms for different tasks in various domains such as Generative AI for text or image analysis, object detection, segmentation, Named Entity Recognition, robotics, ... These different building blocks have historically been developed in different projects, which has led to compartmentalization between projects and few exchanges.

With the arrival of Foundation models for text and vision, it is now crucial to completely manage an MLOps stack shareable across researchers from data management to training and eventually deploy models. The MLOps stack that we are developing ought to perform data management for easy versioning and retrieval. When training, the weights of the resulting machine learning models will be stored in a model registry to keep track of which data, which code version and which hyper-parameters have been used, to ensure reproducibility and good resource management. Lastly, the stack makes validation of the model and possibly its deployment.

The missions of this internsip is to help to develop this MLOps stack that will enable the researchers to easily develop new algorithms from raw data to the deployment and monitoring and Machine Learning Models.



MLOps chain

The intern will be involved in the development of various libraries and frameworks to enable researchers to manage their data through collection to curation and assisted annotation. Tools to train of neural networks on one GPU, on one node or multi-nodes will also be delivered with optimization in mind. The range of training comes from self-supervised learning to fine-tuning and supervised learning depending on the project. Finally, to complete the MLOps chain, models will be deployed internally to perform reliably and efficiently their tasks for our research use. When required, the models are also deployed in the premises or the cloud platform used by the industrial partner.

Also, we'd like to be able to easily deploy our models internally for demonstrations, to use our models to respond to tasks and/or to participate in an annotation phase for new datasets via a dedicated tool.

Internship objectives

During this internship, development can occur in various place of the MLOps stack to :

- Train neural networks efficiently on multiple GPUs and multiple compute nodes for different vision and text tasks such as Generative AI (text and image), object detections, Named Entity Recognition (NER)...
- Deploy models internally to make predictions and evaluate models.
- Monitor our models.
- Store our large quantity of data.
- Maintain a model registry to track our experiments and easily deploy our best performing models.
- visualize and annotate datasets containing multiple modalities (text, image, video, etc.) assisted by AI.

Solutions to orchestrate, maintain and scale the technical infrastructure in our laboratory

You will also perform technological watch of MLOps tools to find the best languages, libraries or frameworks to perform the missions mentioned above.

Qualifications

- Students in their 4th or 5th year of studies (M1, M2 or gap year)
- Python proficiency and preferably in a deep learning framework (especially TensorFlow or PyTorch)
- Skills in Machine learning (deep learning, perception models, generative AI...)
- Knowledge in orchestration (Kubernetes), containerization (Docker) and in Web Development is a plus

Job-related benefits

Joining the CEA List and the LVA as an intern means:

- Joining an organization that addresses societal challenges to build the world of tomorrow.
- Working in one of the most innovative research organizations in the world (ranked in the global top 100, top 3 in France).
- Discovering a rich ecosystem where the institute creates privileged links between the industrial and academic sectors.
- Conducting research in an environment where autonomy and creativity are recognized, and where valorizing results is encouraged (publication of scientific articles, patents, and sharing of open-source code whenever possible).
- Joining a young and dynamic team made up of research engineers, PhD students, post-doctoral researchers, and interns.
- Benefiting from an internal computing infrastructure equipped with around 300 state-of-the-art GPUs.
- Receiving a stipend between €1300 and €1400 per month.
- Having the opportunity to continue with a PhD or as a research engineer after the internship.
- Having the possibility of remote work, receiving a 75% (instead of 50%) reimbursement on public transportation costs, and benefiting from the "mobili-jeune" aid to reduce rent costs...

To apply, contact the laboratory with a CV and cover letter: lva-stages@cea.fr