Tackling Key Challenges of AI Development – Insights from an Industry-Academia Collaboration

Alexander Melde[1]*, Manav Madan[3]*, Paul Gavrikov[2], David Hoof[2], Astrid Laubenheimer[1], Janis Keuper[2], and Christoph Reich[3]

[1] Karlsruhe University of Applied Sciences
{alexander.melde, astrid.laubenheimer}@h-ka.de
[2] Offenburg University
{paul.gavrikov, janis.keuper}@hs-offenburg.de, dhoof@stud.hs-offenburg.de
[3] Furtwangen University
{manav.madan, christoph.reich}@hs-furtwangen.de

Abstract. Harnessing the overall benefits of the latest advancements in artificial intelligence (AI) requires the extensive collaboration of academia and industry. These collaborations promote innovation and growth while enforcing the practical usefulness of newer technologies in real life. The purpose of this article is to outline the challenges faced during cross-collaboration between academia and industry. These challenges are also inspected with the help of an ongoing project titled "Quality Assurance of Machine Learning Applications" (Q-AMeLiA), in which three universities cooperate with five industry partners to make the product risk of AI-based products visible. Further, we discuss the hurdles and the key challenges in machine learning (ML) technology transformation from academia to industry based on robustness, simplicity, and safety. These challenges are an outcome of the lack of common standards, metrics, and missing regulatory considerations when state-of-the-art (SOTA) technology is developed in academia. The use of biased datasets involves ethical concerns that might lead to unfair outcomes when the ML model is deployed in production. The advancement of AI in small and medium sized enterprises (SMEs) requires more in terms of common standardization of concepts rather than algorithm breakthroughs. In this paper, in addition to the general challenges, we also discuss domain specific barriers for five different domains i.e., object detection, hardware benchmarking, continual learning, action recognition, and industrial process automation, and highlight the steps necessary for successfully managing the cross-sectoral collaborations between academia and industry.

Keywords: Artificial Intelligence, Machine Learning Lifecycle, MLOps, Collaboration of Academia and Industry, Challenges in Action Recognition, Model Search

1 Introduction

In recent years, applications of AI in the industry have made significant gains. This highlights a novel trend in ML-based data-driven products which are quickly substituting their traditional counterparts. However, investing in AI product development without contemplating the exact deployment of these products for business value creation can lead to a variety of problems in future which can also be summed up as technical debt [1].

The complexity and vast amount of new technological advancements in AI can quickly become overwhelming for enterprises starting their own AI transition journey. These challenges can be solved with a successful and productive collaboration of universities and industrial partners. When universities and enterprises collaborate to solve complex tasks in AI, the motivational aspects are different for individuals. Enterprises provide real-life input and high-quality data, to refine more generic research into specific problem definitions whereas academia focuses on generating breakthroughs in research. Overall their preferences are not always aligned. This paper highlights some of these differences based on five different use cases.

Additionally, there have been various studies performed in the past outlining the challenges of ML development and deployment. However, most of these are based on conducted interviews with ML experts from different fields and lack comparability [2,3,4]. As ML problems can be highly unique it is only natural that not all challenges can be covered by these different studies.

In this work, first, challenges encountered in collaborations between industry and academia for AI projects will be presented. Following, the challenges related to each use case are highlighted, including ethical issues and explaining general conflicts of interests in this context. Finally, solutions to the most common challenges will be proposed.

* These authors contributed equally.

1.1 Related Work

Despite many different approaches proposed in the past for tackling issues for successful collaboration between the industry and academia, not many include AI as the center of the development.

Some of the studies and their results can be summarized as follows: Aykol et al. [5] mention that tackling challenges in applying ML for the development of complex material systems such as batteries requires successful collaboration between academia and industry. They also refer to the need for experimental battery data which is required for the further growth of research in academia. This is similar to the lack of industry-standard datasets that are observed on a large scale in multiple AI research fields. Additionally, Fursin et al. [6] mention the difficulties of co-designing efficient software and hardware which have emerged in recent times due to the growing technology transfer gap between academia and industry. The use case that is targeted in this work is autotuning, which focuses on automatically exploring optimization spaces such that the efficiency of computer systems could be improved. The key challenges that are highlighted in this work are the lack of a common experimental framework and the lack of practical knowledge exchange between academia and industry. Their viewpoint targets the point that practitioners spend more time adapting already developed tools for their use case of multi-objective autotuning rather than innovating solutions for new problems such as more complex search spaces. They develop a framework and present a study in which researchers and students are taught to use reusable customizable workflows. In their framework, the outputs such as research artifacts can be shared easily with a unified API to further encourage reusability. Furthermore, Garousi et al. [7] target an overall perspective on industry-academia collaboration for the successful development of software engineering. The authors mention three main challenges that should be tackled. These three challenges are (1) conserving and staying on a common goal; (2) engaging in mutual understanding (between all partners) and teamwork, and (3) identifying managerial bottlenecks such that the whole project could be tracked.

None of the studies mentioned above have yet provided a comprehensive overview of the challenges of AI-based products, especially describing the division based on different domains and their dependency on key challenges.

1.2 The Q-AMeLiA project

The motivation behind this paper originates in the project “Quality Assurance for Machine Learning Applications” (Q-AMeLiA). The consortium aims to support small and medium enterprises (SMEs) in the special machine learning software development life cycle (ML-SDLC) by developing tools to evaluate data quality indicators (e.g. in terms of representative coverage of the feature space), as well as to evaluate the quality of the learned AI model achieved in the learning process. Further, to serve the business needs of the industry partners, the consortium aims to reduce and make measurable the product risk of AI-based products, enabling manufacturers to assure a quantified performance of their products (with respect to AI decisions) to their customers (see Fig. 1).

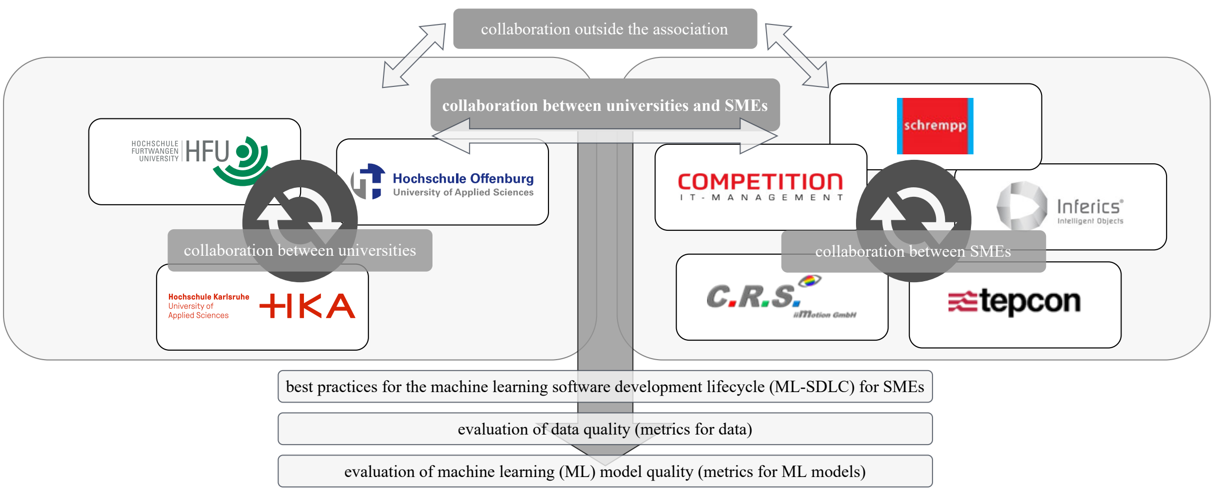

Fig. 1. Consortium: 5 SMEs collaborate with 3 universities.

To achieve these goals, three research topics have been defined. First, universally applicable and domain-independent best practices for the ML-SDLC should be developed by combining research results with use cases and lessons learned from industry. Results regarding this question are shown in this paper and previous work [8]. Second, methods, metrics and tools for qualitative and quantitative assessment of data quality should be developed. An architecture to quantify the risk of AI models has been presented in [9]. Third, these results should be adapted exemplary for a set of real-world use cases provided by the industry partners.

2 Challenges of AI Development

Collaborating on AI products comes with its challenges. Despite the differences in interest, academia and industry have competencies that can enable an interdisciplinary approach to innovation in AI. Indeed, studies have shown that such collaborations have increased in the recent past [10]. This section aims to provide an overview of common challenges faced when collaborating for AI projects.

First, *domain independent* organizational challenges are outlined. These are sometimes non-technical but play an important role in value generation from ML algorithms in the real world. Next, we outline challenges based on five individual domains which we refer to as *domain dependent* challenges.

2.1 Domain-Independent AI Challenges

During this and previous collaborations, the partners of the Q-AMeLiA project encountered common challenges independent from the specific projects domains. In this section, a selection of these challenges will be presented based on both the perceived importance and coverage in related publications [11,12,13,14].

Varying Mindsets In general, academia and industry have disjoint paths to AI. While academia focuses on research and development of new algorithms, seeking to find generalizing solutions for common challenges, industrial interest is tailored to their often highly-specific problems.

There is a vast amount of hype around AI, which in turns causes unrealistic expectations from enterprises. The involvement of multiple stakeholders might further complicate the product as joining business metrics to AI is not a straightforward task. The varying mindsets could result in a conflict of interests and might also lead to compromise on ethical aspects of AI.

Ownership of Intellectual Property From a ML perspective, a product consists of a combination of training data, an algorithm, a model trained on the data, and the output it generates. These all could come from different sources and could be owned by different stakeholders, leading to conflicts regarding the ownership of the final product as it cannot be created without the individual components. Collaborations should therefore discuss intellectual property (IP) early-on and should decide how the IP and usage rights are apportioned throughout the consortium. An example may be whether an enterprise who provides a dataset should be awarded the usage rights of the model trained with the respective data.

Data Quality and Quantity A sufficient amount of high quality training data can be considered a major factor for the success of AI projects and projects fail due to low data quality. Enterprises tend to overlook this aspect due to a target-driven approach, which focuses on generating business value with ML without really understanding the requirements of these approaches. A better approach would be to put the data and the model first [13]. Disregarding this advice may lead to expensive iterations of data-ingestion and curation. AI researcher Andrew Ng proposed that “the focus has to shift from big data to good data” [15]. Hence for any collaboration, the dataset and use case are the first points on which consensus should be reached before any effort is applied.

Dataset Licensing Industry-oriented research requires tailored datasets, as tackling industrial problems based on research-oriented datasets is difficult due to the domain shift. Unfortunately, such datasets are rare because companies are often interested in keeping their datasets closed source, to prevent various privacy issues and to allow for a competitive advantage. To create a representative dataset, the full complexity of a given context should be covered, however, it is impossible to represent the full range of external influence and it is difficult to define where to abstract and simplify. In previous work, data is often recorded in laboratory conditions that are not suitable for real life applications [16], but sometimes even this is a tedious process, as for example recordings of people can be difficult to obtain for privacy reasons. Furthermore, some important events may only occur rarely and therefore be underrepresented in collected datasets. This problem can be seen in anomaly detection systems: Real video recordings capturing all possible events may not be feasible in practice, and on the other hand simulation-based approaches are not guaranteed to behave naturally.

Overchoice in Source Models Transfer learning exploits knowledge learned by a model on one problem to help solve a different but related problem. For this, a model pretrained on a large dataset is fine-tuned on a significantly smaller dataset, considerably cutting data acquisition costs. However, there is a variety of pretrained ML models available for various domains, making it difficult for experts to choose the optimal fit for a particular problem. Furthermore, the optimal model may be different for each problem due to requirements such as latency, performance, and license considerations.

Integration in Legacy Systems Enterprises typically already operate and maintain software solutions. New ML models should then be able to communicate with these legacy systems and the effort required to integrate ML models into them should not be underestimated. Due to fast advancements in ML, there is a requirement for greater interoperability and automation of ML solutions compared to the software stack already in use.

Further Challenges In addition to the challenges faced in collaboration, there are additional barriers that need to be tackled when applying AI to real-life product, e.g. minimizing bias (see section 3), handling legal requirements like copyright and GDPR and choosing the right deployment infrastructure (regarding compute architectures and platforms and energy consumption). In addition, models should be as efficient as possible, explainable (providing the possibility to explore the reasoning behind AI decisions) and sometimes transferable, e.g. from cloud to edge [12].

2.2 Domain-Dependent Challenges

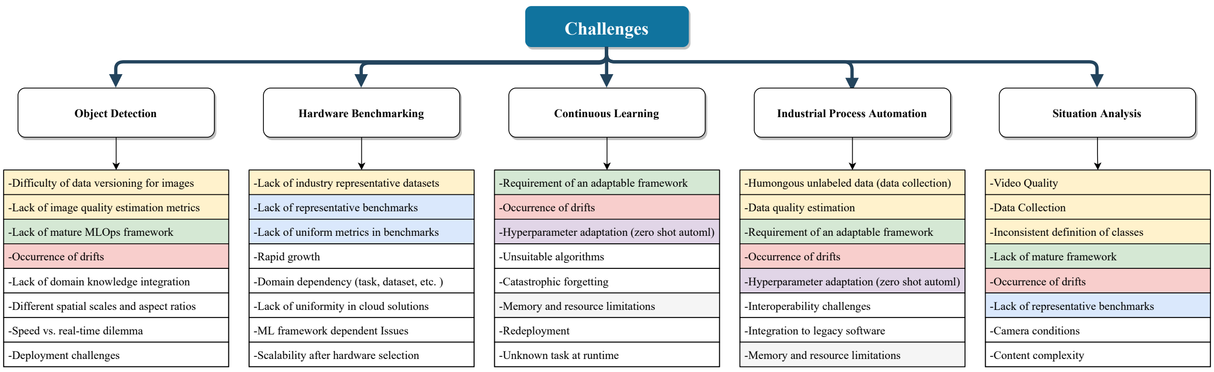

Fig. 2. Overview of challenges for each domain.

In addition to the challenges mentioned in the previous section, some challenges only occur in particular domains. The domains discussed in this paper are chosen due to their relevance in the research community and the importance for our industry partners. A more complete list of challenges can be seen in Fig. 2.

The uniqueness of such challenges arises from either the kind of data required or the use case itself. A total of five different domains of industrial use cases are targeted in the Q-AMeLiA project and are introduced in this section.

Object Detection using ML Q-AMeLiA collaboration partner *C.R.S. iiMotion GmbH* is providing solutions and services in the field of image and video processing. Their current research interest is the construction of an automated object detection pipeline for anomaly detection in optical paths of camera systems.

Anomaly detection is a prominent use case where the identification of defects or frauds is widely implemented in the industry [17]. The aim is to build an automated ML workflow utilizing machine learning operations (MLOps) such that the delivery time of the new model for production can be minimized [8]. This involves building an ML workflow automatically with steps such as automated data quality assessment and model testing.

Hardware Benchmarking for ML Software and hardware solutions for ML are growing at a rapid pace. Unlike other types of computational workloads, ML based workloads vary in compute intensity. A ML workflow is typically divided into two parts: training and inference.

In the training phase, the algorithm is iteratively updated till some predefined metric is reached and hence a lot of resources are required for this process. The use of resource-intensive hardware like graphics processing units (GPUs) and machines accelerated with tensor processing units (TPUs) is crucial for training models faster.

In the inference phase the trained and deployed model is used for making predictions. This is a less resource-intensive task, enabling the deployment to embedded devices with restricted resources.

Choosing the correct machine learning hardware for training is a complicated process. It is known that different workloads might require different specialized hardware [18,19]. Evaluating the suitability of hardware is the Q-AMeLiA use case for the company *competition it-management GmbH* that offers IT consulting services, project management, and support for setting up hardware and software Infrastructure.

Continuous Learning under Drift Traditional ML workloads presume that the data is static and will remain constant also in the production, which is not necessarily the case. The need for Continuous Learning (CL) stems from the fact that data never remains static and neither does the environment. Due to the various drift scenarios, such as data drift, concept drift, and model drift, CL is required to address these challenges and update the workload accordingly. The notion of CL is relatively new in the field of ML. Nevertheless, it is well known that neural networks suffer from the fundamental problem of forgetting what they have learned in the past when trained on new data. This is referred to as catastrophic forgetting. This problem makes the continuous improvement of a static ML workload a difficult task [20]. This is the use case for the company *tepcon GmbH* who encountered these challenges while providing machine learning based digitization solutions for predictive maintenance of industrial machinery.

Industrial Process Automation Regarding industrial process automation, Q-AMeLiA collaboration partner *schrempp edv GmbH* explores intelligent user interfaces (UI) that can predict and provide shortcuts for user interactions from the perspective of enterprise resource planning systems (ERP). ERPs help to track and store different processes and transactions in an organization. However, customer needs may be highly specific and require manual optimization of the UI to speed-up common routines. The rise of data analytics and big data, allows for such systems to become more intelligent such that they can automatically adjust the UI [21] for each individual user. The company collects anonymized telemetry data that it then uses to train predictive ML models. This allows customer-tailored UIs at a fraction of the cost of manual adjustments that can often exceed customers expectations.

Situation Analysis in Health Care Health care can benefit greatly from AI-assisted solutions, as demonstrated by our industry partner *Inferics GmbH*, which provides an AI platform used for access control systems based on face recognition and situation analysis for assisted living based on 3D poses. Detecting swiftly when, for example, elderly people are in dangerous situations and need assistance can be lifesaving, however identifying human posture and activity can be difficult to implement in real use cases for a variety of reasons, which are explained and grouped in more detail below.

For situation analysis, video data is needed for training. It is possible to record this data using varying camera conditions: the camera position influences the scale of an action and the viewpoint [16], possibly leading to (partial) occlusion. In the wild, lighting, shadows and internal camera parameters will vary. In some scenarios, the camera will even move. Varying video quality is encountered in the form of different resolutions, framerates, compression artefacts and blur. In addition, the complexity of the content filmed is in itself high, as a video may contain multiple activities performed by multiple actors either subsequently or in parallel. Every actor has it's own anthropometric appearance [16], just as the background scenes are varying and might distract from the activity. In addition, weather conditions and seasons might change how entire scenes look and how activities are performed. Any action can have varying temporal and spatial boundings, leading to videos (or bounding boxes) that have to be both cutted to time frames and cropped to a certain image area to represent only one action at a time.

Another challenge is the variety in definitions of situations, actions, activities, interactions, events and tasks in previous publications and datasets. These terms are closely related and sometimes used interchangeably. There are attempts to use activities as a general term that can be categorized into other terms based on their complexity, however this approach is not exhaustive [22]. Other work differentiates terms based on the number of humans involved [16]. Even if these definitions are used, an ambiguous term like "running" can be used to describe an action performed by either a single person or a group with diverse contexts and intentions. Some activities involve sub-activites, creating a hierarchy. Both intraclass variety and interclass similarity are high, as tasks can be completed using various methods and some activities are very similar to each other. Due to the potentially infinite number of activities, no dataset can cover the entire spectrum, and no naming convention has yet been established. Existing datasets suffer from label noise and poor label quality caused by online clickers with different cultural backgrounds.

Once the data is collected, high computational complexity is encountered, caused by multi-dimensional data and the involvement of temporal context. Depending on the implementation, high resolution 2d or 3d videos annotated with multi-dimensional labels and additional meta data are used. Unfortunately, the state of the art for situation analysis is far behind the one in e.g. image classification. No de-facto standard libraries, break-through datasets or benchmarks have been established yet. This scattering makes it difficult to compare previous research, whose approaches also vary widely, e.g. in the network architecture used, the temporal context strategy, the feature extraction (separate or end-to-end), the fusion strategy (if required), and the embedding used in the network.

To summarize, there is an insufficient amount of data for training action recognition models (given the high complexity of the task), datasets lack scene variety and their classes are biased, unbalanced and unreliably labeled. Due to the computational complexity and the difficulty in comparing previous approaches, efficient research can only be conducted if sufficient training time and computational resources are available [23].

3 Ethical Issues and Conflicts of Interest

Hagendorff and Meding [24] presented an important study on the ethical impact of the industry in ML research. They mainly set the focus on ethical concerns, conflicts of interest, innovation, and gender equality and conclude the following: (1) even with the growth of collaboration there are rarely any mentions of the conflict of interest between academia and industry, (2) research papers by industry often target trending ML topics earlier than academia on average and (3) the focus of research from the industry often falls short on social aspects such as gender diversity and gender equality.

In this work, first, the ethical concerns regarding AI will be considered, and then, the conflicts of interests between academia and industry will be described separately.

3.1 Ethical AI

Previous research has shown how biased AI software can harm society [25]. Such software might cause damage to under-represented groups in the source data, including discrimination based on race, gender, or otherwise unfair decision-making. With the rising number of AI products, the potentially negative impact of AI may be amplified, creating the need to integrate ethical considerations at the heart of product development. Unfortunately, addressing ethical concerns can result in a loss of business value for the industry and is therefore often skipped. Yet, the effectiveness of laws and regulations tackling ethical issues of AI is limited.

To tackle these problems, the European Commission introduced ethics guidelines for trustworthy artificial intelligence in 2019 [26]. These guidelines specify different principles for general AI ethics but are not an European law with any legal framework. Due to the rapid development of AI, standards are quickly outdated or not common yet. This gap in terms of legal consequences from different regulators leads to a need for self-regulation. The final goal should be to make AI responsible, transparent and accountable, such that the products based on AI could become sustainable in the future.

3.2 Conflict of Interest: Academia vs. Industry

Commercial product development and early-stage research have traditionally been regarded as the major difference between the industry and the academic world. Academic collaboration with industry takes many forms, but specifically for AI, collaboration efforts between the two have not been well coordinated [27]. There is a possibility of conflict between academia and industry in the development and monitoring of the use of AI systems.

In the last decade, companies have hired experienced researchers such as professors from prestigious universities for full-time positions to work on AI on a large scale. This implies that there is a strong motivation for creating applied solutions from the research however in the long run this would amplify the lack of experienced professionals for students at the universities. Among other reasons, researchers are opting out of academia due to the lack of resources. In spite of the above-mentioned challenges, the increased presence of industry could help stimulate the search for innovative solutions to real-world problems, thereby fueling further growth. On the other hand, some in the industry, in general, might strive to develop high-precision systems that may not be developed in accordance with ethical considerations [28]. For example, many of the initial datasets used in building commercial products by different industry giants had exhibited bias in some sense and were still used until publicly criticized [27]. There is also a lack of incentive for companies to act responsibly. They invest far more in AI than academia has done so far and might set their own (profit optimizing) goals higher. A solution to the above challenges lies in successful cooperation between both areas [27].

4 Paths forward

Over the past decade, large enterprises like Netflix, Google, and Amazon have benefited massively from the adoption of ML, but SMEs often fail in their efforts to adopt AI in their businesses. We believe the solution to this problem lies in successful cooperation between academia and industry. The challenges have been categorized in the previous sections as domain-dependent and domain-independent and an overview of domain-dependent challenges is presented in Section 2.2 and Fig. 2. In this section, we further discuss the predominant challenges of chapter 2 and propose solutions based on the collaboration of academia and industry.

4.1 Drift Adaptation

The environment into which a ML model is deployed to can change over time, leading to a change in the models input data and the degradation of a model's performance over time. To counter these so-called data drifts, the deployed model should be updated continuously, e.g. by building an automated MLOps pipeline. MLOps pipelines are closed-loop systems in which either a new model is trained or an old one is retrained, to counter the drop in performance indicated by a specified metric [8]. Drift adaptation should be an internal block of MLOps. In ML, drift is categorized into two parts: covariate drift and concept drift. The ML model is often developed in a static environment, tested, and finally deployed in a dynamic environment. This change can result in to drop in performance as the model might encounter data it has not seen previously or the relationship between the input data and predicted label might also change with time. The former case describes the situation in which there is a change in the distribution of incoming data. This is referred to as the covariate drift. The latter use-case describes the concept drift, where the patterns extracted from the data change over time. Methods to handle concept drift can be divided into two groups: Implicit (blinding) methods update the models in regular intervals, independently of the occurrence of the concept drift, whereas explicit methods update models only if a concept drift occurred [29]. There is a need to establish a general drift adaptation framework that targets all different types of drift.

4.2 Data Collection & Data Quality Estimation

Data is a key aspect of ML because models extract patterns based on the acquired data. Usually, data gets exposed to multiple sources of error at different stages of development. Challenges such as missing data, range miss-match and type miss-match pose specific threat to data integrity. These errors require integrity checks which are placed at different stages of an ML workflow. A standard process of handling data for ML starts with a storage engine (e.g. a warehouse or a data lake) which is then integrated into a feature store. The data from the feature store is then cleaned and transformed and finally integrated into the MLOps pipeline. This end-to-end process from curation to ML model is not standardized and organizations define their own tool stacks based on their requirements. Choosing the right data models and storage engines is challenging due to the high number of options [30][31]. Additionally, data quality is a subjective and a multi-dimensional concept, therefore good quality data cannot be defined with a single standard definition. The quality that differs from context to context is highly dependent on the targeted business process [32].

Another aspect of data preparation for using ML in the computer vision domain is *image labeling*. This step is immensely important as the label quality decides the quality of ML models in supervised learning scenarios. It is common to observe mistakes found in labels of popular benchmark datasets like ImageNet due to sloppiness or imprecise instructions [33]. Given their huge size, such datasets are heavily used in academia for benchmarking and pretraining, however due to labeling mistakes, precautions should be taken when research is applied in real-world settings, creating the need for better datasets. Building new datasets requires labeling which in turn requires domain expertise with multiple iterations for the construction of a high quality dataset. Often, the construction of new datasets is done by individuals with limited domain expertise resulting in non-representative datasets. The goal should be to fill this gap by industry-standard benchmark datasets created by industry experts. The challenges of data labeling are also highlighted in the work by Denton et.al. [34]. They discuss how the quality of labels can be improved with extended feedback loops between the annotators and the experts. Researchers should be incentivized to use more appropriate datasets based on industrial standards for scientific studies. There is a need to provide more encouragement to do more qualitative analysis with results based on the accuracy of the performed tasks. For example, the explainable AI could flow into the results with the interpretability of the trained model. Additional tests could also be carried out to check for fairness and energy efficiency. This is also illustrated by the model cards created in the work by [35]. Furthermore, the academia can help create better metrics for data quality estimation [9].

4.3 Automation and MLOps

Automation aims to boost productivity and make everyday tasks easier. For the ML domain, this is offered by machine learning operations (MLOps) [8]. The emphasis of MLOps is on reducing the friction between the different departments involved in producing business value through the use of ML. It advocates for teamwork and cutting down on waste in terms of the artifacts generated in the ML lifecycle. Many of the challenges with ML arise once the model has been trained and is put into use in production. For example, there are scenarios where the model's performance degrades with time as described in section 4.1 (drifts).

Without MLOps, local manual development workflows are data-scientist-centric. The model and the data both lie with the data scientists who do the training until they transfer the model to the operations department responsible for deployment. This static workflow has a lot of challenges, can introduce many different errors and vulnerabilities and deployed models cannot be easily replaced with new ones.

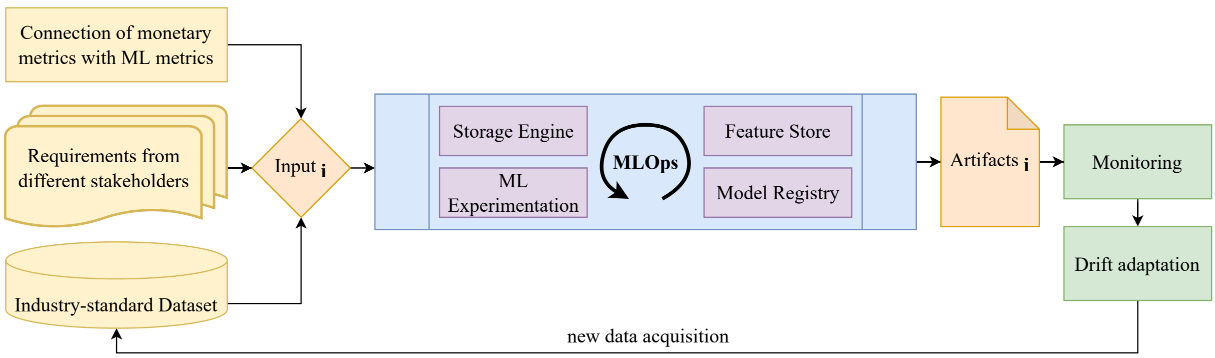

Fig. 3. Overview of an MLOps workflow.

MLOps provides a set of tested guidelines for reliably and automatically resolving such issues. An exemplar workflow is portrayed in Fig. 3. From the experience gathered through this project, it is important to understand that there is not one single correct MLOps workflow. There could be additional blocks for different processes in MLOps as this is not standardized for every domain. But the important part is to build the workflow with the option of integrating input from different stakeholders with industrial standard datasets to reduce the time spent on experimentation on the non-representative datasets. The workflow displayed also shows that the process of monitoring can be connected with an intelligent drift detection block to keep the quality of the ML model deployed high. There should also be additional support for data acquisition for the next iteration of the development. Through this, the ML models could be continuously adapted in the future automatically.

The MLOps approach to building a machine learning lifecycle focuses on automation and scalability. It provides the necessary stack of tools for rapid deployment of new models and helps in versioning and logging the success of deployments, creating a governance loop.

However, it can be challenging to know where to start and how to adopt MLOps practices. For businesses that are not performing the same operations manually, quantifying the benefits of MLOps is another challenge. For this reason, a survey of open-sourced tools was conducted, laying the foundation for adopting MLOps depending on the maturity of data science processes in a company [8]. The survey [8] serves as an introduction to various actors, roles, and tools involved in MLOps workflows. It provides a thorough comparison of supportive tools for the different phases to simplify the process of MLOps adoption. After MLOps has been integrated, a system for continuous learning that is capable of retrieving fresh training data, creating new features, developing new models, testing, and registering can easily be discovered.

4.4 Q-AMeLiA Search Engine

Due to the lack of high quantities of high-quality training data, ML engineers oftentimes opt for already trained models as a starting point for training a model on the new dataset. This process is referred to as transfer learning which typically yields considerable performance gains, especially for cases where data is scarce, however, in computer vision, it is specifically difficult as the data is of a high dimension and training from scratch can be an expensive process. For different domains, numerous pretrained models are available publicly, but choosing the best one is not straightforward.

As a solution to these problems regarding finding the right model, a search engine for pretrained computer vision models was developed. It helps to look up models by leveraging user-provided basic meta information such as the training task, visual domain, architecture, input size, or number of parameters. This first version of the search engine is available open-source under the following URL:

`https://github.com/Q-AMeLiA/searchengine`

Ongoing research is conducted to automate the process of model selection, and, a second version of the search engine is planned, which will suggest pretrained models based on a small, user-provided data subset, and a backend-sided evaluation of model suitability based on multiple metrics such as quality of the learned representations [36,37].

5 Conclusion

Usually, industry experts and researchers from academia have different backgrounds and objectives. With the ongoing project Q-AMeLiA it is showcased how bottlenecks in different domains of AI can be identified by merging the interests of academia and industry. The project has resulted in approaches for (1) linking theoretical research with creative solutions, (2) capturing challenges of individual domains, and (3) highlighting the bottlenecks in the automation of ML workflows for continual improvement in production. In the future the project will focus on creating further solutions with the help of the industrial expertise of the partners involved. In general, such collaborations encourage innovative solutions to real-world problems. When universities are linked to the business world, they are inundated with problems requiring solutions which in feedback provides new research directions to the universities. This is also helpful for the universities in the sense of curriculum design as the project highlighted the gap between applied machine learning (ML) and its research. The knowledge learned can be used to train students how to tackle the challenges of applying ML at scale in the real world. Students working on the project were also able to understand the relevance of both theoretical and practical approaches with the aid of such collaboration.

References

1. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.F., Dennison, D.: Hidden technical debt in machine learning systems. Advances in neural information processing systems **28** (2015)
2. Baier, L., Jöhren, F., Seebacher, S.: Challenges in the deployment and operation of machine learning in practice. In: ECIS. (2019)
3. Paleyes, A., Urma, R.G., Lawrence, N.D.: Challenges in deploying machine learning: a survey of case studies. arXiv preprint arXiv:2011.09926 (2020)
4. Lwakatare, L.E., Raj, A., Bosch, J., Olsson, H.H., Crnkovic, I.: A taxonomy of software engineering challenges for machine learning systems: An empirical investigation. In: International Conference on Agile Software Development, Springer, Cham (2019) 227–243
5. Aykol, M., Herring, P., Anapolsky, A.: Machine learning for continuous innovation in battery technologies. Nature Reviews Materials **5**(10) (2020) 725–727
6. Fursin, G., Lokhmotov, A., Savenko, D., Upton, E.: A collective knowledge workflow for collaborative research into multi-objective autotuning and machine learning techniques. arXiv preprint arXiv:1801.08024 (2018)
7. Garousi, V., Felderer, M., Fernandes, J.M., Pfahl, D., Mäntylä, M.V.: Industry-academia collaborations in software engineering: An empirical analysis of challenges, patterns and anti-patterns in research projects. Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE) (2017)
8. Ruf, P., Madan, M., Reich, C., Ould-Abdeslam, D.: Demystifying mlops and presenting a recipe for the selection of open-source tools. Applied Sciences **11**(19) (2021)
9. Melde, A., Laubenheimer, A., Link, N., Schauer, C.: An architecture to quantify the risk of ai-models. Upper-Rhine Artificial Intelligence Symposium UR-AI 2021 (2021) 84–93
10. Michel-Schneider, U.: Challenges for University - Industry Collaboration - a Stakeholder View. Proceedings of Business and Management Conferences 12713398, International Institute of Social and Economic Sciences (October 2021)

11. Mikhaylov, S.J., Esteve, M., Campion, A.: Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. Philosophical transactions of the royal society a: mathematical, physical and engineering sciences **376**(2128) (2018) 20170357
12. German AI Association (KI-Bundesverband): Large European AI Models (LEAM) als Leuchtturmprojekt für Europa (Konzeptpapier). LEAM:AI (6 2022)
13. Nahar, N., Zhou, S., Lewis, G., Kästner, C.: Collaboration challenges in building ml-enabled systems: Communication, documentation, engineering, and process. Organization **1**(2) (2022) 3
14. Saghiri, A.M., Vahidipour, S.M., Jabbarpour, M.R., Sookhak, M., Forestiero, A.: A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions. Applied Sciences **12**(8) (2022) 4054
15. Strickland, E.: Andrew ng: Unbiggen ai. IEEE Spectrum (2022)
16. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video Technology **18**(11) (2008) 1473–1488
17. Tao, X., Zhang, D., Ma, W., Liu, X., Xu, D.: Automatic metallic surface defect detection and recognition with convolutional neural networks. Applied Sciences **8**(9) (2018) 1575
18. Mattson, P., Cheng, C., Diamos, G., Coleman, C., Micikevicius, P., Patterson, D., Tang, H., Wei, G.Y., Bailis, P., Bittorf, V., Brooks, D., Chen, D., Dutta, D., Gupta, U., Hazelwood, K., Hock, A., Huang, X., Kang, D., Kanter, D., Kumar, N., Liao, J., Narayanan, D., Oguntebi, T., Pekhimenko, G., Pentecost, L., Janapa Reddi, V., Robie, T., St John, T., Wu, C.J., Xu, L., Young, C., Zaharia, M.: Mlperf training benchmark. In Dhillon, I., Papailiopoulos, D., Sze, V., eds.: Proceedings of Machine Learning and Systems. Volume 2. (2020) 336–349
19. Madan, M., Reich, C.: Comparison of benchmarks for machine learning cloud infrastructures. CLOUD COMPUTING 2021 (2021) 50
20. van de Ven, G.M., Tolias, A.S.: Three scenarios for continual learning. ArXiv **abs/1904.07734** (2019)
21. Tallón-Ballesteros, A.: The design of erp intelligent sales management system. Fuzzy Systems and Data Mining VI: Proceedings of FSDM 2020 **331**(2020) (2020) 413
22. Vrigkas, M., Nikou, C., Kakadiaris, I.A.: A review of human activity recognition methods. Frontiers in Robotics and AI **2** (2015)
23. Gowda, S.N., Rohrbach, M., Sevilla-Lara, L.: Smart frame selection for action recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. Volume 35/2. (2021) 1451–1459
24. Hagendorff, T., Meding, K.: Ethical considerations and statistical analysis of industry involvement in machine learning research. AI & SOCIETY (2021) 1–11
25. Giuffrida, I.: Liability for ai decision-making: some legal and ethical considerations. Fordham L. Rev. **88** (2019) 439
26. Larsson, S.: Ai in the eu: Ethical guidelines as a governance tool. In: The European Union and the Technology Shift. Springer (2021) 85–111
27. Horvitz, E.: One hundred year study on artificial intelligence (2016)
28. Lauer, D.: You cannot have ai ethics without ethics. AI and Ethics **1**(1) (2021) 21–25
29. Mehmood, H., Kostakos, P., Cortes, M., Anagnostopoulos, T., Pirttikangas, S., Gilman, E.: Concept drift adaptation techniques in distributed environment for real-world data streams. Smart Cities **4**(1) (2021) 349–371
30. Huyen, C.: Designing Machine Learning Systems. " O'Reilly Media, Inc." (2022)
31. Idreos, S., Callaghan, M.: Key-value storage engines. In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. SIGMOD '20, New York, NY, USA, Association for Computing Machinery (2020) 2667–2672
32. Panahy, P.H.S., Sidi, F., Affendey, L.S., Jabar, M.A.: The impact of data quality dimensions on business process improvement. In: 2014 4th World Congress on Information and Communication Technologies (WICT 2014), IEEE (2014) 70–73
33. Vasudevan, V., Caine, B., Gontijo-Lopes, R., Fridovich-Keil, S., Roelofs, R.: When does dough become a bagel? analyzing the remaining mistakes on imagenet. arXiv preprint arXiv:2205.04596 (2022)
34. Denton, E., Díaz, M., Kivlichan, I., Prabhakaran, V., Rosen, R.: Whose ground truth? accounting for individual and collective identities underlying dataset annotation. arXiv preprint arXiv:2112.04554 (2021)
35. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T.: Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. (2019) 220–229
36. Gavrikov, P., Keuper, J.: Cnn filter db: An empirical investigation of trained convolutional filters. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (June 2022) 19066–19076
37. Gavrikov, P., Keuper, J.: Adversarial robustness through the lens of convolutional filters. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. (June 2022) 139–147

Acknowledgements

We thank our industry partners for helpful comments and discussions. This work was funded by the Ministry of Science, Research and Arts of Baden-Württemberg (MWK) as part of the project "Quality Assurance of Machine Learning Applications" (Q-AMeLiA).