Towards practicable Machine Learning development using AI Engineering Blueprints

Nicolas Weeger
Ansbach UAS
Ansbach, Germany
nicolas.weeger@hs-ansbach.de

Annika Stiehl
Ansbach UAS
Ansbach, Germany
annika.stiehl@hs-ansbach.de

Jóakim von Kistowski
Aschaffenburg UAS
Aschaffenburg, Germany
joakim.vonkistowski@th-ab.de

Stefan Geißelsöder
Ansbach UAS
Ansbach, Germany
stefan.geisselsoeder@hs-ansbach.de

Christian Uhl
Ansbach UAS
Ansbach, Germany
christian.uhl@hs-ansbach.de

***Abstract*—The implementation of artificial intelligence (AI) in business applications holds considerable promise for significant improvements. The development of AI systems is becoming increasingly complex, thereby underscoring the growing importance of AI engineering and MLOps techniques. Small and medium-sized enterprises (SMEs) face considerable challenges when implementing AI in their products or processes. These enterprises often lack the necessary resources and expertise to develop, deploy, and operate AI systems that are tailored to address their specific problems.**

Given the lack of studies on the application of AI engineering practices, particularly in the context of SMEs, this paper proposes a research plan designed to develop blueprints for the creation of proprietary machine learning (ML) models using AI engineering and MLOps practices. These blueprints enable SMEs to develop, deploy, and operate AI systems by providing reference architectures and suitable automation approaches for different types of ML.

The efficacy of the blueprints is assessed through their application to a series of field projects. This process gives rise to further requirements and additional development loops for the purpose of generalization. The benefits of using the blueprints for organizations are demonstrated by observing the process of developing ML models and by conducting interviews with the developers.

***Index Terms*—Machine Learning, AI Engineering, Blueprints, Reference Architecture, MLOps**

I. Introduction

Artificial intelligence (AI) is transforming numerous industries and application domains. It is crucial for organizations to adopt AI techniques in order to achieve business success [1], [2]. SMEs in particular can benefit significantly from adopting AI techniques. They can improve or even establish capabilities in certain areas, such as customer experience, production monitoring, and decision-making processes [3].

The proprietary development of ML models for use as or within a product, called AI systems, in an organizational context can lead to a number of challenges [4]–[7]. This includes understanding of the intricacies of AI, including its functional requirements and operational scenarios. The integration of additional processes, such as data generation and preprocessing or model training and deployment, with the traditional software engineering process can potentially create organizational constraints, particularly for SMEs [6]. The ML model development lifecycle includes additional practices beyond DevOps for the data and models. These include MLOps and DataOps techniques that provide a culture, practices, and tools for handling data and models. In addition, the system architecture for AI systems must be aligned with the requirements of the underlying model. Training and inference environments, along with data storage and versioning of the different artifacts, must be incorporated in order to enable the system to function as an AI system.

Consequently, the effective implementation of AI systems requires a well-designed architecture that is tailored to the specific requirements of the intended AI application. This paper discusses the importance of AI engineering practices and the concept of developing blueprints tailored to the requirements of different types of AI and stages of development.

We will discuss the content of these blueprints and how they can be utilized by organizations. The discussion will commence with a preliminary result in the form of a pipeline and its sub-pipelines.

The results of this research will assist organizations to address these challenges and streamline the development, deployment and operation of AI systems. Consequently, they will be enabled to adopt the blueprints by providing reference architectures and suitable automation approaches for various types of AI.

II. Background and Related Work

A. AI Engineering

AI engineering is an evolution of the field of software engineering and, due to the rapid growth of ML developments, it is an emerging field. AI engineering is currently at the maximum "peak of inflated expectations", as evidenced by Gartner's AI Hype Cycle for 2024[1]. According to Gartner, "AI

[1]https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence

2768-4288/25/$31.00 ©2025 IEEE
DOI 10.1109/ICSA-C65153.2025.00078

engineering is the foundation for enterprise delivery of AI and GenAI at scale. Most organizations lack the data, analytics and software foundations to move individual AI projects to production at scale - much less operate a portfolio of AI solutions at scale." Over a dozen projects were examined in [8], where AI engineering challenges led to problems with productizing ML models. The study found that the majority of companies that develop machine learning models encounter difficulties when trying to move them into production. They provide a research agenda and overview of the issues that need to be addressed in this direction. [9] point out that while the field of software engineering research has been extensively discussed, AI engineering has been much less addressed. Only a limited number of publications present concrete experiences related to the application of AI engineering principles. They selected ten AI engineering practices in several categories from the literature and applied them to an example implementation to evaluate the practices and their systems architecture.

Furthermore, discussions and questionnaires, especially with SMEs, have shown that there is a desire to implement AI in their systems. However, the success of their model implementation and productization depends on overcoming the challenges mentioned above. Thus, applying AI engineering practices can help organizations streamline the development, deployment and operation of machine learning models.

B. MLOps

The idea of MLOps is to provide techniques and tools for the deployment and operation of AI systems [10]. The goal is to devise a strategy for solving real-world problems with the deployment of ML models. Several studies review various literature in this area and offer pipelines, taxonomies, tools, methods and challenges in this area [11]–[13].

[14] conduct a systematic mapping study for MLOps architectures and point out 35 architecture components, describe different architecture variants for different use cases and provide popular tools to use for these architecture components.

C. Other Related Work

In [15] a reference architecture for the specific use cases in the process industry dealing with edge devices was presented. They demonstrated the architecture by implementing a case study for a real-world use case and proved the functionality with this application.

[16] developed a reference architecture to facilitate the use case of big data in edge computing ML techniques. They come up with different views on the architecture of model development and deployment for this specific use case.

Another study [17] presents a vision for "disciplined, repeatable and transparent model-driven development and Machine-Learning operations (MLOps) of intelligent enterprise applications." They provide a three-stage metamodel for model based development of AI/ML blueprints based on intelligent application architecture.

With scope on software and architecture, design patterns for AI based systems are discussed in several studies [18]–[21]. They provide an overview of design patterns, adapted or newly generated for AI use cases, and show the application and the resulting benefits of using them to develop machine learning models.

In summary, the literature provides insights into the importance, potential architectures, and principles for AI engineering and MLOps practices. However, the application of these insights is currently focused on a few reference architectures in specific domains, such as big data or edge devices. Other studies focus on defining architectures and patterns and prove their applicability in case studies.

The principles of AI engineering provide the foundation for the development of the blueprints proposed in this paper. MLOps pipelines and tools as well as existing reference architectures and frameworks are employed to support the development of AI systems, thereby streamlining, standardizing, and accelerating the process. Software and architecture design patterns are used to describe the development in order to fulfill the non-functional requirements (NFRs) for the different blueprints. The application in field projects enables flexible, highly automated deployment and resource-efficient operation for different requirements in SMEs.

III. Methodology

Since the literature is sparse when it comes to practical examples of AI engineering directly usable for SMEs, this work is intended to develop blueprints for common AI use cases. The blueprints comprise the description of suitable reference architectures and reference applications, automation pipelines, and tools for model development, deployment, and operation.

The extant literature offers a variety of interpretations of MLOps pipelines, which encompass the principles of DataOps and DevOps as they relate to ML [11]–[13]. Previous considerations of these MLOps pipelines have led to the formulation of a guiding pipeline for the development of blueprints (Fig. 1).

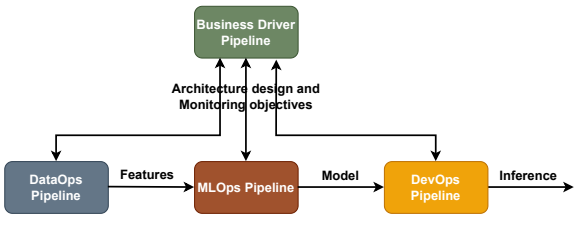

Fig. 1. Guiding pipeline with four sub-pipelines.

First, the business requirements are defined and the architectural components are specified. This serves as a framework for implementing and validating all subsequent stages. The DataOps pipeline prepares the data for model training in the MLOps pipeline. The resulting model artifact is then incorporated into software using the DevOps pipeline to enable inference.

The following subsections provide a detailed breakdown of these sub-pipelines. They elaborate the methodology for developing blueprints for each of these pipelines, for different types of AI, and deployment scenarios. The AI types, including computer vision, time series analysis, reinforcement learning, and generative AI, as well as the manifold deployment scenarios possess systematically different requirements that lead to varied architecture definitions and thus the necessity for multiple blueprints.

The combination of these blueprints according to the requirements of a project will facilitate a standardized and simplified approach that will enhance the efficiency of model development and operation.

A. Business driver pipeline

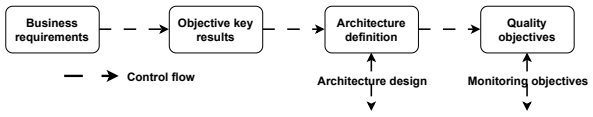

Fig. 2. Business driver pipeline for defining requirements, architecture and objectives.

The business driver pipeline (Fig. 2) provides a methodology for defining requirements, examining measurable objective key results (OKRs), and defining architecture for the software and ML components as well as quality objectives for the monitoring. The business requirements are then discussed with the organizational stakeholders and subsequently combined with the NFRs of the corresponding AI type and translated into measurable and testable OKRs. Subsequently, these are employed to specify the architectural framework for the software components, including the different levels of design patterns and tools to be used for the pipeline architecture, in addition to the model architecture and data structure alternatives. Objectives for monitoring the model performance, environment utilization and data quality are defined to allow the generation of monitoring alerts in production scenarios.

B. DataOps pipeline

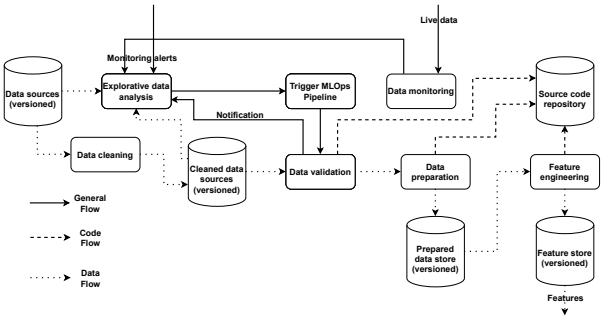

Fig. 3. DataOps pipeline to prepare the data for model training.

The DataOps pipeline (Fig. 3) includes all stages of data processing for model training. This covers the full range of activities, from exploratory data analysis and data cleaning to data validation, preparation and feature engineering. A key aspect of the DataOps pipeline is the versioning of the data. This enables the reproducibility of each stage of data manipulation. Note that not all use cases will require all of these steps. The blueprints should include options for tools that can be adapted to different use cases, as well as predefined steps that can be used for data processing with different requirements.

C. MLOps pipeline

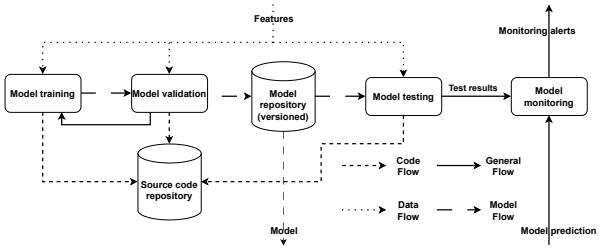

Fig. 4. MLOps pipeline for model training, validation and versioning.

The MLOps pipeline (Fig. 4) incorporates components for model training, validation, and testing. These components utilize features from the feature store that was created by the DataOps pipeline. The blueprints for this phase include information about the tools to be employed for the automation of the training process, the tracking of experiments, and the storage of versioned model artifacts for the experiments. Furthermore, the blueprints should identify different training environments that facilitate the implementation of high-performance training for different types of ML and data.

D. DevOps pipeline

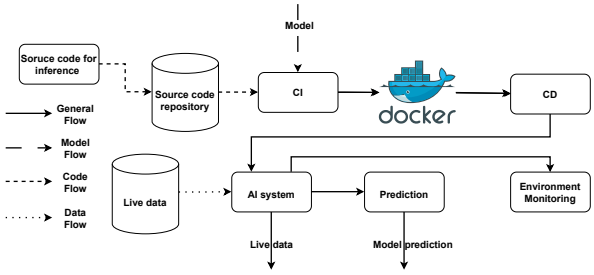

Fig. 5. DevOps pipeline for system integration of the model artifact.

The DevOps pipeline (Fig. 5) covers the system integration of the final model artifact, using continuous integration (CI) and continuous deployment (CD) for the prediction on live data in an AI system. This includes source code to enable inference, such as batch and API interfaces, and the packaging or

containerizing of the usable model service for productization. The blueprints include options for interface implementation and different deployment strategies, such as cloud, edge or standalone deployment, depending on the requirements of the use case. Monitoring of model performance, environment utilization, and live data quality is also defined and discussed.

E. Research method

In order to examine the blueprints and validate their usability for enterprises, the design science research (DSR) method [22], [23], shall be employed. By means of interviews and a comprehensive review of relevant literature, the challenges and requirements of SMEs are identified, and potential avenues for improvement are ascertained.

Based on these findings, business requirements can be established in collaboration with the relevant stakeholders, aligning with their specific needs. By integrating the business requirements with the requirements of the various types of AI being utilized, for instance, algorithms, data storage, computational demands, and NFRs, a comprehensive framework can be devised to facilitate the verification of the designed artifacts. Subsequently, these can be subjected to iterative testing and validation. Finally, the artifacts can be deployed in the stakeholders' projects as a field test. This process is then to be repeated until the requirements have been finalized and the artifacts have been demonstrated to fulfill them. The process is repeated in multiple projects to generalize the findings, making them as applicable as possible for SMEs.

IV. Conclusion

This paper shows the need for research that helps small and medium sized enterprises to integrate development of ML models into their organizations. To address the challenges inherent in this process, the proposal suggests the use of blueprints that include reference architectures, pipelines, and tools for model development, deployment, and operation, in conjunction with reference applications tailored to the specific requirements of different types of AI.

It is essential that the blueprints be as specific as possible with regard to a particular type of AI and deployment strategy, while also being sufficiently general to be beneficial to a wide range of applications. To truly enable the streamlined application of AI for SMEs, we argue that requirements from different projects must be successively integrated into generalized blueprints. In turn, these blueprints must be tested against multiple use-cases in real world applications.

References

[1] I. M. Enholm, E. Papagiannidis, P. Mikalef, and J. Krogstie, "Artificial Intelligence and Business Value: A Literature Review," *Information Systems Frontiers*, vol. 24, no. 5, pp. 1709–1734, 2022.

[2] S. M. C. Loureiro, J. Guerreiro, and I. Tussyadiah, "Artificial intelligence in business: State of the art and future research agenda," *Journal of Business Research*, vol. 129, pp. 911–926, 2021.

[3] K. Bhalerao, "A study of barriers and benefits of artificial intelligence adoption in small and medium enterprise," *Academy of Marketing Studies Journal*, vol. 26, no. 1, 2022.

[4] L. Fischer, L. Ehrlinger, V. Geist, R. Ramler, F. Sobiezky, W. Zellinger, D. Brunner, M. Kumar, and B. Moser, "AI System Engineering—Key Challenges and Lessons Learned," *Machine Learning and Knowledge Extraction*, vol. 3, no. 1, pp. 56–83, 2020.

[5] L. E. Lwakatare, I. Crnkovic, and J. Bosch, "DevOps for AI – Challenges in Development of AI-enabled Applications," in *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. Split, Croatia: IEEE, 2020, pp. 1–6.

[6] M. Schönberger, "Artificial Intelligence for Small and Medium-sized Enterprises: Identifying Key Applications and Challenges," *Journal of Business Management*, vol. 21, pp. 89–112, 2023.

[7] E. D. S. Nascimento, I. Ahmed, E. Oliveira, M. P. Palheta, I. Steinmacher, and T. Conte, "Understanding Development Process of Machine Learning Systems: Challenges and Solutions," in *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. Porto de Galinhas, Recife, Brazil: IEEE, Sep. 2019, pp. 1–6.

[8] J. Bosch, H. H. Olsson, and I. Crnkovic, "Engineering AI Systems: A Research Agenda," in *Advances in Systems Analysis, Software Engineering, and High Performance Computing*, A. K. Luhach and A. Elçi, Eds. IGI Global, 2021, pp. 1–19.

[9] M. Grote and J. Bogner, "A Case Study on AI Engineering Practices: Developing an Autonomous Stock Trading System," in *2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)*, 2023, pp. 145–157.

[10] G. Symeonidis, E. Nerantzis, A. Kazakis, and G. A. Papakostas, "MLOps - Definitions, Tools and Challenges," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. Las Vegas, NV, USA: IEEE, 2022, pp. 0453–0460.

[11] M. Testi, M. Ballabio, E. Frontoni, G. Iannello, S. Moccia, P. Soda, and G. Vessio, "MLOps: A Taxonomy and a Methodology," *IEEE Access*, vol. 10, pp. 63 606–63 618, 2022.

[12] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," *IEEE Access*, vol. 11, pp. 31 866–31 879, 2023.

[13] M. Steidl, M. Felderer, and R. Ramler, "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice," *Journal of Systems and Software*, vol. 199, 2023.

[14] F. A. Najafabadi, J. Bogner, I. Gerostathopoulos, and P. Lago, "An Analysis of MLOps Architectures: A Systematic Mapping Study," in *European Conference on Software Architecture*, vol. 14889. Springer Nature Switzerland, 2024, pp. 69–85.

[15] R. Wostmann, P. Schlunder, F. Temme, R. Klinkenberg, J. Kimberger, A. Spichtinger, M. Goldhacker, and J. Deuse, "Conception of a Reference Architecture for Machine Learning in the Process Industry," in *2020 IEEE International Conference on Big Data (Big Data)*. Atlanta, GA, USA: IEEE, 2020, pp. 1726–1735.

[16] P. Pääkkönen and D. Pakkala, "Extending reference architecture of big data systems towards machine learning in edge computing environments," *Journal of Big Data*, vol. 7, no. 1, pp. 1–29, 2020.

[17] W.-J. Van Den Heuvel and D. A. Tamburri, "Model-Driven ML-Ops for Intelligent Enterprise Applications: Vision, Approaches and Challenges," in *Business Modeling and Software Design*, B. Shishkov, Ed. Cham: Springer International Publishing, 2020, vol. 391, pp. 169–181.

[18] L. Heiland, M. Hauser, and J. Bogner, "Design Patterns for AI-based Systems: A Multivocal Literature Review and Pattern Repository," in *2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)*, 2023, pp. 184–196.

[19] R. Sharma and K. Davuluri, "Design patterns for Machine Learning Applications," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. Erode, India: IEEE, 2019, pp. 818–821.

[20] R. Cabral, M. Kalinowski, M. T. Baldassarre, H. Villamizar, T. Escovedo, and H. Lopes, "Investigating the Impact of SOLID Design Principles on Machine Learning Code Understanding," in *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI*. Lisbon Portugal: ACM, 2024, pp. 7–17.

[21] M. Take, S. Alpers, C. Becker, C. Schreiber, and A. Oberweis, "Software Design Patterns for AI-Systems," *EMISA*, pp. 30–35, 2021.

[22] Hevner, March, Park, and Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, p. 75, 2004.

[23] M. Ivarsson and T. Gorschek, "A method for evaluating rigor and industrial relevance of technology evaluations," *Empirical Software Engineering*, vol. 16, no. 3, pp. 365–395, 2011.