Markus Böhm, Jürgen Wunderlich (Hrsg.): AKWI Jahrestagung 2024,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2024 79

Harmonising innovation and governance: A lifecycle model for high-risk AI systems under the European AI Act

Full Paper

Carolin Gotsch[1], Jörg Puchan[2]

Abstract: The rapid advancement of Artificial Intelligence (AI) technologies has sparked a discussion within organisations, questioning whether regulatory frameworks like the European Artificial Intelligence Act (EU AI Act) pose obstacles or opportunities for innovation and growth. In response to this ongoing discourse, this paper introduces a comprehensive lifecycle model tailored for high-risk AI systems. On the basis of a literature review, state-of-the-art Machine Learning (ML)/AI and software development lifecycles were identified to establish a foundational framework. By examining the requirements of the AI Act, specific to high-risk AI systems, actionable steps were extracted and integrated into the lifecycle. The resulting framework was developed iteratively, incorporating adaptations from identified lifecycles and mapping essential compliance steps. Expert interviews provided valuable insights for refinement, leading to a universal and future-proof lifecycle. The proposed framework not only ensures compliance with regulatory standards but also fosters innovation and development in the AI landscape.

Keywords: Artificial Intelligence; AI Lifecycle; Reference Model; AI Act; High-Risk AI Systems

1 Introduction

While Artificial Intelligence (AI) systems provide a lot of opportunities regarding efficiency and productivity to businesses, experts from the field also warn about the potential risks of the technology and call for a regulation. [KS23; An23] In order to address these threats, the European (EU) Parliament and the EU member states reached an agreement on the European Artificial Intelligence Act (EU AI Act) on the 9th December 2023, which is the first ever comprehensive regulation on AI. [Vo23; Eu23a] The aim of the act is to ensure AI in Europe being safe and respecting fundamental rights, while also thriving innovation and the expansion of businesses. [Eu24] Therefore, certain applications are banned by the law, whereas others are subject to specific obligations based on their potential risk. [Eu24] Regulatory sandboxes and real-world testing should foster innovation and are especially promoted to small and medium enterprises enabling development and training of AI before placing it on the market.

[1] Hochschule München, Fakultät für Wirtschaftsingenieurwesen, Lothstraße 64, 80335 München, caro.gotsch@gmail.com
[2] Hochschule München, Fakultät für Wirtschaftsingenieurwesen, Lothstraße 64, 80335 München, puchan@hm.edu

doi:10.18420/AKWI2024-006

[Eu23a] A new institution, the so-called AI Office, has already been established within the EU on 21st February 2024 and will play a key role in implementing the regulation by developing tools, methodologies and benchmarks. [Eu24] After the entry into force, which is likely to be by end of May 2024, the act becomes fully applicable after a transition period of two years, with certain provisions coming into effect sooner (AI Act, Art. 85). As fines for non-compliance will be up to €35 million or 7% of the companies' turnover, whichever is higher [Eu23a], it can be seen that the impact of non-compliance with the AI Act may have a severe impact on providers or deployers of AI systems. In addition, there are already several opinions on the question of whether or not the act will lead to overregulation and thus to an innovation brake. [KNS23] This discussion also motivates the underlying work of this paper, the aim of which was to examine the impact of the EU AI Act on the development of AI systems. The following sections therefore provide an understanding of the theoretical background (chapter 2) and the methodology (chapter 3). Afterwards, an AI systems lifecycle, which is compliant to the AI Act, is presented and serves as the research questions solution in form of a reference model (chapter 4). Finally, the work is concluded and limitations are revealed (chapter 5).

2 Background

2.1 AI Act

For the purpose of this paper, the most recent version of the AI Act is used, which is the text of the provisional agreement, dated on 2nd February 2024. Within this paper it is often referred to the Annexes II, III and VII of the regulation, consequently this term always indicates a reference to the AI Act. When citing specific parts of the law, the related article of the AI Act is specified through the abbreviation "Art." in parentheses.

The act applies to AI systems placed on the European market, put into service or used in the EU. Therefore, obligations of providers, deployers and other parties, such as distributors and importers are in the scope of the act (Art. 2). In the abovementioned version of the act an AI system is defined as "[…] a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" (Art. 3 (1)). However, not the technology itself is regulated by the act but the application that the AI system serves. Therefore, a risk-based approach categorises them into four risk levels: AI systems, that pose an "unacceptable risk", a "high risk", a "limited risk" and a "minimal or no risk".

Practices that are in scope with AI systems with an unacceptable risk are prohibited, for instance: real-time remote biometric identification, social scoring and emotion recognition in the workplace and educational institutions (Art. 5).

The focus of the law is on the regulation of high-risk systems as most of the text addresses the requirements for this risk category. There are two ways in which AI systems can fall into this risk category. Firstly, the system is used as a product or as a safety component across a product spectrum including civil aviation, vehicle security, marine equipment, toys, lifts, pressure equipment and personal protective equipment (Annex III). Furthermore, it must be subject to third-party conformity assessment in accordance with that Annex II legislation (Art. 6(1)). Secondly, the system is automatically classified as high-risk if it falls within one of the following uses cases: non-banned biometrics, critical infrastructure, education and vocational training, employment and workers management and access to self-employment, access to and enjoyment of essential public and private services, law enforcement, migration and asylum and border control management, administration of justice and democratic processes (Annex III; Art. 6(2)). However, exceptions apply to systems within these use-cases, which are not likely to lead to a significant risk to the health, safety or fundamental rights and may circumvent the requirements of the high-risk category. Additionally, AI systems profiling individuals are always considered high-risk (Art. 6(2a)). The majority of obligations fall on providers of high-risk AI systems for example the establishment of a risk and quality management system, conduction of data governance, design of a record keeping functionality and implementation of human oversight (Title III, Chapter 2 of the AI Act). All requirements are described in detail in chapter 4 of this paper.

Lastly, AI systems, which interact with humans, like chatbots, emotion recognition and biometric categorisation systems and AI systems that generate or manipulate image, audio or video content (deep fakes) are in the scope of limited risk systems. They underly obligations for the purpose of informing the user about him interacting with an AI system and in case of AI generated content it must be disclosed the automated creation of the content. Further, a voluntary code of conduct for none high-risk systems can be created (Art. 52). The risk of AI systems, which do not fall under one of the aforementioned applications is classified as minimal or none-existent and the systems are not subject to intervention by the AI Act. So-called general-purpose AI (GPAI) models show a significant level of generality and are designed to perform a wide range of tasks across different domains rather than being specialized in one specific area. [Art. 2(44b); Tr24] Providers of such GPAI models need to cooperate with the high-risk AI systems providers to enable compliance. Furthermore, they are obliged to create technical documentation on the model's training, testing and evaluation. Downstream providers are supplied with information regarding model integration and copyright compliance (Art. 52c). Systemic GPAI models, identified by the cumulative amount of compute used for its training being greater than 10^25 floating point operations, (Art. 52a) are required to be notified to the Commission by their providers. They need to undergo additional measures, such as adversarial testing, risk assessment, incident tracking, and ensuring cybersecurity protection (Art. 52d).

2.2 State of the art ML/AI lifecycles and software development lifecycles (SDLCs)

In order to derive requirements for the design of the problem's solution, the state of the art regarding established SDLCs, AI/ML lifecycles and process models in data mining projects was reviewed. Through key term search as well as reference mining of relevant sources, the following models were identified as a suitable base for this work.

First of all, the CRISP-DM (Cross Industry Standard for Data Mining), may be seen as the most common industry-independent methodology for data mining, analytics, and data science projects even more than twenty years after its release. [SKM21] It is divided into six sequential phases, starting with business understanding, followed by data understanding, data preparation, modelling, evaluation and finishing of with the deployment phase. [Ch00; WH00] During the literature review it has been identified, that the development of ML applications is mainly based on these phases. However, as the CRISP-DM is limited in managing requirements of current technologies [ADP18], further models were taken into account. More flexible methodologies have been introduced by big IT companies [Am19], for instance the TDSP (Team Data Science Process) by Microsoft, which has a more agile and iterative approach in order to deliver predictive analytics solutions. [Ma21] Continuous integration (CI) and continuous delivery (CD) paradigm is the current industry standard for automating and streamlining the process of building, testing and deploying code changes. [La22; KAA20] Thus, the integration of DevOps as one of the mostly used agile processes is integrated into the solution delivered by this paper. The DevOps process focuses on collaboration between development and operation teams and includes the steps: plan, code, build, test, release, deploy, operate, monitor and feedback. [KAA20; LCB20]

3 Methodology

The research objective has been elaborated with the Design Science Research (DSR) approach, which is a research paradigm that focuses on the development and validation of prescriptive knowledge in an iterative way. [Pe06] Due to its applicability to application-oriented problem solving in the intersection of IT and organization [Pe06], it is suitable for the research question of this work. The DSR process includes six steps in order to develop and evaluate a so-called artefact. [Pe07]

Problem identification and motivation

As already pointed out, the requirements of the AI Act are considered as extensive, especially for high-risk AI systems. Also, the formulation of these requirements in the act is rather theoretical. Thus, there is a need to investigate the explicit impact of the upcoming legislation across the lifecycle of a high-risk AI system.

Objectives of the solution

The solution process requires the generation of an artefact, which in this case represents the lifecycle of these systems. It must include the functional requirement of assuring compliance to the AI Act whilst developing and operating the system efficiently. As the features of the state of the art SDLCs and ML lifecycles are well established in practice, they were considered as structural requirements for the developed artefact. This includes the AI Act compliant lifecycle to be divided into specific phases with belonging activities, to ensure an agile approach through feedback loops between phases and to be application-neutral.

Design and development of the artefact

Based on the objectives, the artefact has been developed combining the deductive and inductive approach for creating a reference model. [Fe14] Therefore, constructs from generally accepted frameworks, already presented in 2.2, have been adopted to map the necessary process steps across the lifecycle of an AI system. Afterwards, the requirements of the AI Act for high-risk AI systems were extracted from the latest draft of the act and implemented into this lifecycle in order to ensure the AI systems compliance to the regulation.

The complete lifecycle is presented within chapter 4. However, an example of how the AI Acts requirements have been mapped onto the lifecycle is given at this point for reference. Therefore, the requirement of Data and Data Governance derived from Art. 10 of the AI Act has been chosen: Mapped mainly in the Data Management and Machine Learning Workflow sections, the requirements for robust data governance practices are placed where data-related processes are dominant. This includes data acquisition, preprocessing, and model training stages. Ensuring data relevance, representativeness, freedom from errors, completeness, and addressing biases are crucial during these stages to build reliable and fair AI systems. Processing personal data, the compliance with GDPR must be ensured. It was assumed, that data preprocessing plays a crucial role in achieving a model with a performance as high as possible. Recently the realisation arose, that in many AI projects, the leverage for improving model performance lies in the curation of the training data used, which is also known as data-centric AI. [Ja24] Hence, the data and data governance requirements were not mapped as an additional activity, as the text in the law is rather vague and refers to the provider acting appropriately regarding the intended purpose. Furthermore, bias mitigation measures must take place during the operation of the system, when the model is retrained with new data. The mapping here emphasizes the need for thorough handling and scrutiny of data, which is fundamental to the training and validation of AI models. The underlying master thesis of this paper points out how the mapping of the remaining requirements has been conducted.

Through the back iteration by the activity “demonstration and evaluation” the deductively developed generic lifecycle was inductively revised processing the empiric findings from this phase.

Demonstration and evaluation

In this work the steps demonstration and evaluation have been combined as the underlying regulatory framework was not in force at the time of this work and the model may not have been directly demonstrated in specific use cases or experiments. However, with the growing popularity of AI, the need of AI governance has found its way into science as well as into the industry, even before the upcoming legislation. [La22] Therefore, interviews have been conducted with experts from the field of AI development evaluating the developed artefact on the one hand and on the other hand gaining valuable insights into AI governance practices the companies already have in place. Six interviewees representing both companies with more than 200.000 employees and start-ups counting less than 50 employees have been selected. This ensures, that potential biases regarding the subjective perception of the AI Acts impact on the organisation were identified and balanced. They were selected from the interviewer's environment and through referrals from previous interviewees. As the topic requires both expertise in AI development and familiarity with the regulation, the profiles of the informants differ in order to cover both areas.

Qualitative data was collected through semi-structured interviews in order to adapt the questions, which arose during the interview. [Mi19] The developed framework of the AI Act compliant lifecycle is structured into six major phases, that serve as thematic complexes [Mi19], each of which was presented individually and feedback was then obtained from the interviewees. As semi structured interviews are characterized by a certain openness, the analysis requires a method that makes it possible to address both predetermined and spontaneous themes and statements. [BLM14] The qualitative content analysis allows an open exploration of the data without having fixed hypotheses or assumptions in advance and is therefore a suitable method in achieving the goal of processing the valuable insights of the interviewees. [BLM14] The transcripts were categorised by marking meaningful statements, segmenting the data into significant units and classifying them into categories. [GL10; Mi19; Ma15] While the interviews were conducted according to the structure of the six major phases of the developed lifecycle, the initial categorisation of the data followed this structure as well. During the encoding of the data, subcategories under each of the lifecycle phases arose iteratively. [GL10]

The findings regarding the second phase of the lifecycle are allocated for reference in the following and are illustrated by Figure 1 Analog to the mapping of the AI Acts' requirements, the complete data is analysed in detail in the underlying thesis, which also provides the transcripts of the interviews. The informants brought up that the availability of suitable data resources is a key prerequisite for ML and AI projects. However, as this phase requires a relatively high effort one of the interviewees mentioned, that they conduct a proof-of-concept prior running through all the data management process steps. The informant described their practice as follows: "Simply have the data made available to us, as large data sets as possible, so that we can throw them into […] different models to see if we get meaningful results." (Interviewee 1) All of the experts addressed the topic of bias in the data. Bias mitigation measures should be explicitly included in the

lifecycle. One interview partner provided insights on how their practices look like: “we do this on a correlation basis. What we analyse is how the performance correlates with different parameters of the model. And then we try to correct those biases in the model performance.” (Interviewee 6) At the same time, it was emphasised that the database must be as trans-parent as possible, even when it comes to mitigating bias. The negative example of Googles AI model ‘Gemini’ was mentioned, which failed in generating historically accurate pictures of people in an attempt at racial and gender diversity. In terms of data security, the im-portance of data ownership and security of data must be considered and a reference to GDPR could be included into the lifecycle, as this regulation applies at this point and must be com-plied to especially when processing personal data.

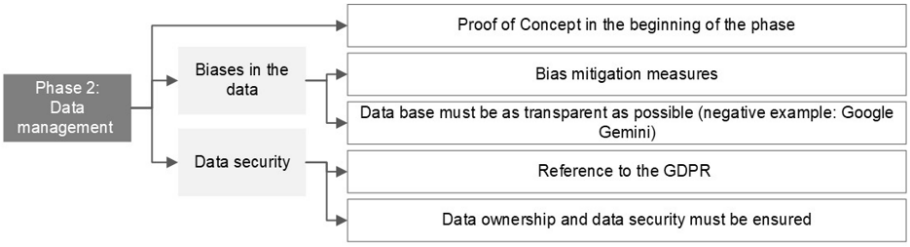

Figure 1: Results of the data analysis categorising the interview data (white boxes) according to the lifecycle phase (grey box)

With the findings from the interviews the artefacts quality was improved by iterating back to the design and development phase. [Pe07]

Communication

The problems’ relevance and its solution in the form of an AI Act compliant lifecycle for AI Systems is published through this paper and its underlying master thesis.

4 Development of an AI Act compliant lifecycle for AI systems

In the following sections, the reference model of an AI Act compliant lifecycle for AI systems is presented. It consists of five phases: business understanding (4.1), data management (4.2), ML/AI modelling (4.3), application development (4.4), deployment and operation (4.5). After each phase is explained individually, their interactions and phase agnostic parts of the lifecycle are discussed (4.6). The model language Business Process Modelling Notation (BPMN) was chosen, as it features comprehensibility for all relevant stakeholders with different backgrounds. [Ke13]

4.1 Phase 1: business understanding

The business understanding phase, illustrated in Figure 2, is initiated by the definition of the use case, followed by identifying and analysing stakeholder needs, defining the targets and assessing whether or not and how ML or AI technologies is suitable for the use case. [Am19] These activities may be run through iteratively. However, at some point it must be checked, if the system will be in scope of the AI Act and subject to its obligations. Therefore, the regional scope of the EU market must be given. Furthermore, the acts' definition of an AI system must include the use case (Art. 3(1)). If the system is not in scope of the act, it does not need to be compliant and is not in scope of this work. As the act follows a risk-based approach, a risk classification needs to be conducted to find out, which obligations of the act apply. Firstly, it needs to be clarified, whether the system is in scope of the GPAI models definition (Art. 2(44b)) and it must be differentiated between GPAI models posing a systemic risk (Art. 52a) and those that do not. However, neither the development of GPAI systems nor their use is in scope of this work.

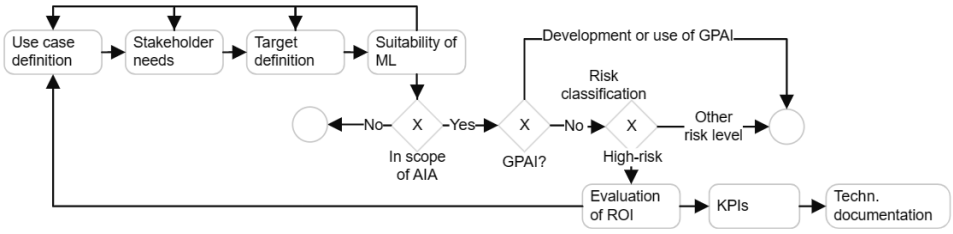

Figure 2: Phase 1 of the lifecycle - Business understanding

Subsequently, the risk classification is performed, which initiates by confirming, that the system will not pose an unacceptable risk and will therefore be prohibited (Art. 5). It is verified if the system is falling under one of the two possible paths of being classified as a high-risk system (Art. 6). Firstly, it is investigated whether the system falls under the Union Harmonisation Legislation listed in Annex II. Fulfilling this first condition, it is checked if a third-party conformity assessment is necessary under the specific legislation (Art. 6(1)). With these two checks being positive, the system is classified as high-risk. In cases, where the first condition is not fulfilled, it needs to be checked, if the system belongs to the use cases listed in Annex III (Art. 6(2)). Additionally, there must be a significant risk (Art. 6(2a)) in order that the system will become a high-risk system. Next, the costs of the system including costs of compliance and operating costs to estimate the return on invest (ROI) need to be evaluated. At this point, back-iteration is also possible, as the cost estimations could indicate non-profitability of the use case. After defining key performance indicators (KPIs) of the system, the first part of the technical documentation required by the AI Act (Art. 9) is created. This includes information regarding: functions, purposes, target groups, application contexts and performance features of the system.

4.2 Phase 2: data management

As data preprocessing is crucial for the efficiency and accuracy of training AI or ML models, and due its complexity and time consuming nature [Ta24; Kr16; BB01], a proof of concept may demonstrate the feasibility of the use case before starting with the activities in this phase, visualized in Figure 3. During the data acquisition, internal or external data sources are identified and the data is collected. [Ch00] Initial explorative analyses help in understanding the data and discover possible problems. [Ch00; SM21] Depending on the use case the AI Act provides exceptions, which permit the usage of particular categories of personal data, for instance to eliminate bias (Art. 10(5)). Therefore, it needs to be checked if such data is acquired to ensure compliance with the General Data Protection Regulation (GDPR). Before the data preparation, quality requirements must be defined, which the data sets need to fulfil. The data preparation step includes multiple activities, which are highly dependent on the use case and on the model that they should be trained on. Therefore, the entire second phase is also strongly connected to the third phase (see 4.3), where the modelling takes place. However, data cleansing, identification and elimination of bias, data labelling, feature selection and data splitting into training, test and validation data sets are necessary steps during data preparation with this list not being complete. [Am19; SM21] Providers are obliged to establish "data governance and management practices appropriate for the intended purpose of the AI system" (Art. 10). Data quality is controlled in accordance with the predefined quality requirements and in case of failure, back iteration to data preparation and data acquisition is possible. If the data quality requirements have been met, data versioning enables to track changes made to the data over time. Finally, the technical documentation regarding data sources and processing, description of the data and data governance practices is created (Art. 11).

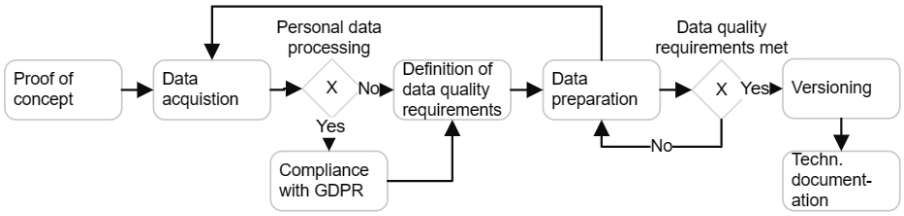

Figure 3: Phase 2 of the lifecycle - Data management

4.3 Phase 3: ML/AI modelling

This phase, featured in Figure 4, starts by the selection of ML/AI model techniques, which need to fit to the problem statement of the use case. [LCB20] Additionally, this activity also has an impact on making the underlying model of the system explainable. [Pa23] It can be summarized, that the act focuses in a holistic manner on ensuring an appropriate use of the system through a transparency requirement being achieved by the

provision of relevant documentation (Art. 13). However, techniques for explainable AI (XAI) are not explicitly addressed by the act but may facilitate human oversight (Art. 14), and the transparency requirement (Art. 13). [Pa23; Ar20] Therefore, at this point of the lifecycle the model could be developed interpretable, so that inner wordings are available and understandable to humans. This could be achieved by a Transparency by Design approach [Fe20], meaning, amongst others, the avoidance of black box models by rather choosing model types, of which the decision-making process is replicable and easy to step through, e.g. decision trees and linear and case-based reasoning models. [Pa23; Fe20] Depending on the use case and the required model type, this is not always possible though.

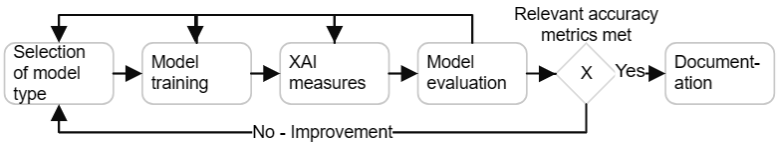

Figure 4: Phase 3 of the lifecycle – ML/AI modelling

The next activity is the training of the model with the training data set including the tuning of hyperparameters to achieve the highest possible output on the validation data set. After the training, the XAI approach may provide techniques in order to make the decision-making process explainable to humans overseeing the system. [Pa23] In contrast to the interpretable AI approach, this approach includes techniques, that may be model agnostic and allow explanation for black-box models. [Pa23] However, this so-called post-hoc explainability comes with its challenges and limitations and is subject to complex research [Pa23; Ar20]. During the model evaluation, the performance using the test data set is measured by testing against metrics such as accuracy, precision, recall and F-score. [LCB20; SFR23] As an "appropriate level of accuracy, robustness, and cybersecurity" of the system is required (Art. 15), the testing against the accuracy metrics needs to be included in the model evaluation. Further information on benchmarks and measurement methodologies addressing these metrics, are announced to be developed in cooperation with relevant stakeholders and organisations but are not yet available (Art. 15). Based on identified deficiencies and patterns, the model is improved for example by adapting feature engineering, parameters or the model selection. [SFR23] This highly iterative process may be stepped through multiple times until the model evaluation is satisfying and is then concluded by the documentation of this phase. [Am19; Ch00; SM21]

4.4 Phase 4: application development

The ML/AI model developed and trained in the previous phase is deployed and somehow serves an application or system, the development of which takes place in this phase, see Figure 5. [SM21]

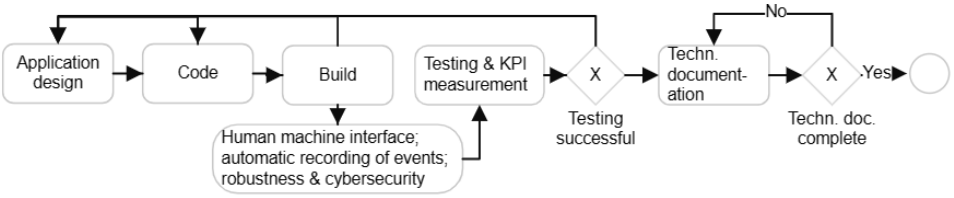

Figure 5: Phase 4 of the lifecycle - Application development

Even though, the use case has already been defined previously and designs may exist through conducting the proof of concept, the application design activity initiates this phase in order to ensure an agile way of working. During the coding and build activities, the functionality of the software is created and functionalities of all components should be orchestrated in such a way that the trained model operates correctly with the rest of the system. [LCB20; SFR23] Within these activities, various requirements of the AI Act must be implemented. Ensuring the system can be overseen at any time during its use by a natural person, who can also intervene to prevent and minimize risks, human oversight measures need to be implemented (Art. 14) Either these measures are built in before placement on the market or it is guaranteed, that the implementation is appropriate to be done by the user. Furthermore, the automatic recording of events (Art. 12) is crucial in order to facilitate post-market monitoring and to identify potential risks. Therefore, especially for high-risk systems, which are under the use cases of Annex III paragraph 1, point (a), the logs should include time of the systems use, input data as well as the database against which these data has been checked and, in some cases, even further information. As already mentioned, accuracy, cybersecurity, resilience and robustness are also requirements to be considered and appropriate measures need to be implemented (Art. 15). Robustness in AI systems means, that they are resilient against adversarial attacks and that they are able to withstand perturbations in input data and model parameter, so that the likelihood of systems failure is reduced [OE19, Fe24]. Suitable measures for achieving robustness depend on the type of the model and system and need to be comprehensive. Literature describes solutions such as adversarial training, which involves training neural networks against adversarial examples [Fe24; Ma17]. However, navigating the techniques and strategies of robustness may be seen as a research topic in itself, but it should be mentioned, that at this point in the lifecycle, it may be necessary to iterate back to the modelling phase (4.3) in order to be able to implement appropriate measures accordingly. Prior to the testing of the system, planning of the tests and definition of test cases must take place in order to determine the systems functionality effectively [FK23], for instance, by conducting integration tests and acceptance testing [FK23]. With regard to software testing, it is distinguished between manual and automated tests. Automated tests are an important part of CI/CD and may be integrated into the agile DevOps workflow [FK23], though, the testing process and tools need to be appropriate for the specific use case. Analogous to the accuracy (chapter 4.3), benchmarks and measurement methodologies for robustness levels are being developed by the Commission and will be decisive for the testing activity soon (Art. 15). With

finalizing the development of the application, several parts of the technical documentation need to be completed (Art. 11), including information regarding components required by the AI Act, that were implemented in this phase. A detailed checklist of the information required in the technical documentation is provided in the work on which this paper is based, hence why it is not discussed in detail here. Furthermore, instructions for use (Art. 13) need to be created with the purpose of not only serving the user of the system but also ensure compliance across the value chain.

4.5 Phase 5: deployment and operation of the system

Before placement on the market the high-risk AI system must undergo a conformity assessment (Art. 43). Depending on the use case of the system, it is distinguished between different types of conformity assessments, which are summarized in Table 1. As Figure 6 illustrates, the CE marking (Art. 49) is affixed and an EU declaration of conformity (Art. 48) must declare the conformity with Chapter 2 of the act, which comprises all the AI Act requirements mapped across the lifecycle. Systems that are within the use cases of Annex III (see 2.1) must be registered in an EU database (Art. 51). High-risk systems with the exception of use cases in scope of critical infrastructure (Annex III point 2) need to be registered at national level. Finally, the system is placed on the market. At this point responsibilities are transferred from the provider to the roles of importers, distributers, and deployers. The presented lifecycle pays primary attention on the obligations of providers. In the underlying work of this paper requirements for the other roles are outlined to make the lifecycle applicable for further participants across the value chain.

Type of system (use case)	**Type of conformity assessment**
1.: Annex III **points 2-8; point 1** (only if **harmonised standards** or **common specifications** are **applied** (Art. 40&41))	Internal conformity assessment (Annex VI)
2.: Annex III point 1 use cases, where **no/only parts of harmonised standards or specifications** are available/were applied	Conformity assessment with the involvement of a notified body (Annex VII)
3.: System under Union **harmonisation legislation** (Annex II) part A	Conformity assessment under specific act with inclusion of Chapter 2 (AI Act) requirements and Annex VII, point 4.3, 4.4, 4.5, and 4.6§5

Table 1: Types of conformity assessments under the AI Act

While operating the system, recording of the logs should take place, which are essential for model monitoring and risk management. Also, generated data and collection of feedback is processed by iterating back into previous phases through feedback loops, which is further discussed in 4.6.

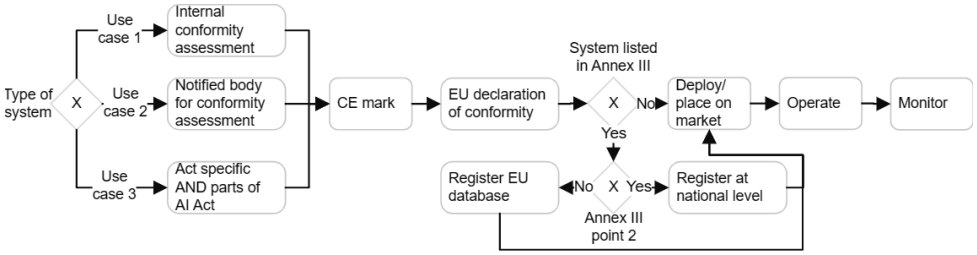

Figure 6: Phase 5 of the lifecycle - Deployment and operation

4.6 General parts of the lifecycle

Interactions in every direction between the single phases of the lifecycle ensure proper transfer of data, other components of the system and feedback. [LCB20; SFR23] This is visualised in Figure 7. It fosters an agile working and development style, which is state of the art, as pointed out in chapter 2.2. If, for instance, data is generated during the operation of the system (phase 5), which is used for model training (phase 3) in order to adapt the model to changing conditions, the data needs to go through various steps of the data management (phase 3). In this case measures against possible bias must be taken (Art. 15). Depending on the impact of the feedback and potential change while iterating backwards, obviously some process steps may be redundant and can be skipped. However, it should always be considered, how a change may influence other components of the system and which further adaptions might be required, especially with regard to compliance with the AI Act. Across the entire lifecycle of the high-risk AI system a risk management system is being operated (Art. 9). It forms the backbone of requirements mentioned above. [PZ24] Especially in terms of the obligations regarding transparency, data governance, accuracy, robustness and cybersecurity, there may be competing priorities, which force the provider to make compromises in reaching the highest possible levels in each of them [Pa23]. Therefore, providers must analyse potential risks, that are reasonable and foreseeable as well as ones, that may arise based on the post market monitoring data (Art. 9). Subsequently, risk management measures are being developed in accordance with a three step-approach of “eliminate or reduce”, “mitigate and control”, and “inform” (Art. 9(4)). The European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENEC) do currently work on developing standards [Eu23b] in order to provide practical guidance for manufacturers by translating the regulation into actionable steps. [PZ24] However, as an already existing standard the ISO/IEC 23894 serves as a guidance on managing risks of AI systems [PZ24] and the National Institute of Standards and Technology (U.S) (NIST) released an Artificial Intelligence Risk Management Framework (AI RMF 1.0). Following this framework, the risk management is split into four phases “govern”, “map”, “measure” and “manage”, each of which entailing specific actions and outcomes. [Ta23]

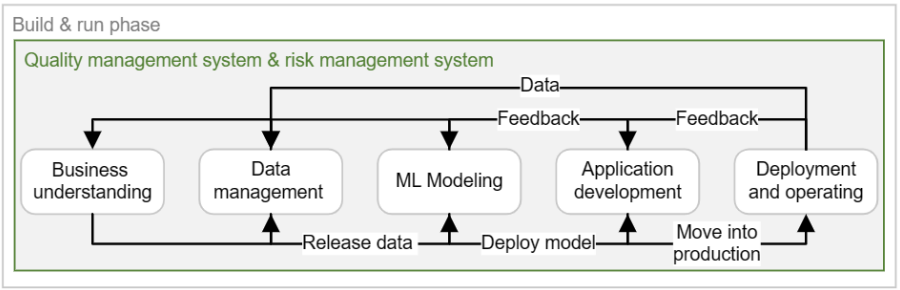

Figure 7: High-level view of the lifecycle

5 Limitations and conclusion

In this paper the impact of the AI Act on the lifecycle of high-risk AI system was analysed by developing a reference model, which is application and sector agnostic. It was focused in particular on the requirements set out in Title III, Chapter 2 of the AI Act affecting mainly the provider of the system. Starting the development of the model from existing state of the art models for ML/AI lifecycle and software development, it could be found, that the existing lifecycles must be revised in order to ensure compliance with the regulation. Although, providers and further participants of the value chain do have up to 2 years of time until the regulation becomes fully applicable, some actions within in the presented lifecycle may already be taken now in order to be prepared, for instance to manage risks of the AI system. Conducting the interviews with experts from the field, it was noticeable, that there are mixed opinions on the impact of the AI Act especially regarding competitiveness with international competitors not being affected by the EU regulation and workload, which may impact time-to-market negatively. Further research could therefore analyse the gap between start-ups and larger companies. However, in order to do this, the development of standards and more concrete interpretations of the AI Act may be necessary. This also had a limiting effect on this work and may be balanced by iterating the developed reference model again as soon as more tangible information concerning the AI Act do exist. Furthermore, deep dives in each of the lifecycle's steps are possible and necessary, as the existing model focuses on the whole lifecycle and is scraping the surface of further profound research topics. The decommissioning phase of the high-risk AI system is also not considered within this work and could be part of further research.

In conclusion, the insights presented in this paper may foster future discussion on the implementation of the AI Acts requirements into the lifecycle of AI systems, leading to the creation of best practices for ensuring compliance with the regulation.

6 Bibliography

[ADP18] Ahmed, B.; Dannhauser, T.; Philip, N.: A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects: 2018 10th Computer Science and Electronic Engineering (CEEC), pp. 11–14, 2018.

[Am19] Amershi, S. et al.: Software Engineering for Machine Learning: A Case Study: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). IEEE, pp. 291–300, 2019.

[An23] Anger, H. et al.: Warum KI-Unternehmen strengere Gesetze fordern und die Politik nicht liefert. Handelsblatt, 2023.

[Ar20] Arrieta, A. B. et al.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, pp. 82–115, 2020.

[BB01] Banko, M.; Brill, E.: Scaling to very very large corpora for natural language disambiguation: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, USA, pp. 26–33, 2001.

[BLM14] Bogner, A.; Littig, B.; Menz, W.: Interviews mit Experten. Eine praxisorien-tierte Einführung. Springer Fachmedien Wiesbaden, pp.71-75, 2014.

[Ch00] Chapman, P. et al.: CRISP-DM 1.0 Step-by-step data mining guide, 2000.

[Eu23a] Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI, 2023.

[Eu23b] European Commission: Implementing Decision on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence. C(2023)3215, 2023.

[Eu24] European Commission: European AI Office, 2024.

[Fe14] Fettke, P.: Eine Methode zur induktiven Entwicklung von Referenzmodellen. In (Kundisch, D. Ed.): Tagungsband Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014). 26.-28. Februar 2014 in Paderborn. Univ, Paderborn, 2014.

[Fe20] Felzmann, H. et al.: Towards Transparency by Design for Artificial Intelligence. Science and Engineering Ethics 26, 2020.

[Fe24] Ferrara, E.: The Butterfly Effect in artificial intelligence systems: Implications for AI bias and fairness. Machine Learning with Applications 15, p. 100525, 2024.

[FK23] Forgács, I.; Kovács, A.: Modern software testing techniques. A practical guide for developers and testers. Apress, New York, NY, 2023.

[GL10] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. Lehrbuch. VS Verlag, Wiesbaden, pp. 198-215, 2010.

[Ja24] Jakubik, J. et al.: Data-Centric Artificial Intelligence. Business & Information Systems Engineering, 2024.

[KAA20] Karamitsos, I.; Albarhami, S.; Apostolopoulos, C.: Applying DevOps Practices of Continuous Automation for Machine Learning. Information 7/11, p. 363, 2020.

[Ke13] Kelemen, Z. D. et al.: Selecting a Process Modeling Language for Process Based Unification of Multiple Standards and Models, 2013.

[KNS23] Klöckner, J.; Neuerer, D.; Scheppe, M.: Unternehmen wollen mit KI radikal umsteuern. Handelsblatt, 2023.

[Kr16] Krishnan, S. et al.: ActiveClean: An Interactive Data Cleaning Framework For Modern Machine Learning: Proceedings of the 2016 International Conference on Management of Data. Association for Computing Machinery, New York, NY, USA, pp. 2117–2120, 2016.

[KS23] Koch, M.; Scheuer, S.: Der Frankenstein-Moment: Wenn wir Künstliche Intelligenz nicht kontrollieren, kontrolliert sie uns. Handelsblatt, 2023.

[La22] Laato, S. et al.: AI governance in the system development life cycle: Crnkovic (Hg.) 2022 – Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI, pp. 113–123.

[LCB20] Lwakatare, L. E.; Crnkovic, I.; Bosch, J.: DevOps for AI – Challenges in Development of AI-enabled Applications: 2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). IEEE, pp. 1–6, 2020.

[Ma15] Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz, Weinheim, Basel, 2015.

[Ma17] Madry, A. et al.: Towards Deep Learning Models Resistant to Adversarial Attacks, 2017.

[Ma21] Martínez-Plumed, F. et al.: CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering 8/33, pp. 3048–3061, 2021.

[Mi19] Misoch, S.: Qualitative Interviews. De Gruyter Oldenbourg, Berlin, Boston, 2019.

[OE19] OECD: Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/00449, 2019.

[Pa23] Panigutti, C. et al.: The role of explainable AI in the context of the AI Act: 2023 ACM Conference on Fairness, Accountability, and Transparency. ACM, New York, NY, USA, pp. 1139–1150, 2023.

[Pe06] Peffers, K. et al.: The design science research process: A model for producing and presenting information systems research. Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST, 2006.

[Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 3/24, pp. 45–77, 2007.

[PZ24] Pouget, H.; Zuhdi, R.: AI and Product Safety Standards Under the EU AI Act, CARNEGIE Endwoment for International Peace, 2024.

[SFR23] Steidl, M.; Felderer, M.; Ramler, R.: The pipeline for the continuous development of artificial intelligence models—Current state of research and practice. Journal of Systems and Software 199, p. 111615, 2023.

[SKM21] Schröer, C.; Kruse, F.; Marx Gómez, J.: A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science 181, pp. 526–534, 2021.

[SM21] Schreckenberg, F.; Moroff, N. U.: Developing a maturity-based workflow for the implementation of ML-applications using the example of a demand forecast. Procedia Manufacturing 54, pp. 31–38, 2021.

[Ta23] Tabassi, E.: Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST Trustworthy and Responsible AI, National Institute of Standards and Technology (U.S.), Gaithersburg, MD, 2023.

[Ta24] Tawakuli, A. et al.: Time-series data preprocessing: A survey and an empirical analysis. Journal of Engineering Research, 2024.

[Tr24] Triguero, I. et al.: General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance. Information Fusion 103, p. 102135, 2024.

[Vo23] Volkery, C. et al.: EU beschließt umfangreichstes KI-Gesetz der Welt – das sind die wichtigsten Punkte. Handelsblatt, 2023.

[WH00] Wirth, R.; Hipp, J. Eds.: CRISP-DM: Towards a Standard Process Model for Data Mining, 2000.

European Parliament: PROVISIONAL AGREEMENT RESULTING FROM INTERINSTITUTIONAL NEGOTIATIONS Subject: Proposal for a regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts 2021/0106(COD) (COM(2021)0206 – C9-0146(2021) – 2021/0106(COD)); 02.02.2024