Linking Team-level and Organization-level Governance in MLOps through XAI and RAI Connector

Elie Neghawi
Electrical and Computer Engineering
Concordia University
Montreal, Canada
e_negh@encs.concordia.ca

Jun Huang
Electrical and Computer Engineering
Concordia University
Montreal, Canada
jun.huang@concordia.ca

Zerui Wang
Electrical and Computer Engineering
Concordia University
Montreal, Canada
zerui.wang@concordia.ca

Yan Liu
Electrical and Computer Engineering
Concordia University
Montreal, Canada
yan.liu@concordia.ca

Abstract—Adoption of AI systems has been widely used across multiple industry domains at an alerting rate without the focus on it's ethical concerns. To address those concerns, there are an increase number of AI ethics frameworks that have been suggested recently that focus on the algorithmic level rather on the systems level. Nonetheless, some of the system level approaches developed mostly cover a single level governance pattern of the system components in the entire software design lifecycle. However, the need to go beyond the single level system design AI ethics frameworks to allow not only a better responsible-AI-by-design, but also a trustworthy process patterns that abstract and link the underlying layers of responsible AI on each and every level. This paper illustrates a principal-to-practice guide of the multi-level governance within organizations across the globe for AI ethics frameworks. We outline the main areas of gap in organizations for AI ethics frameworks. Consecutively, we propose a multi-level governance pattern for responsible AI systems within organizations which is participatory, iterative, flexible and operationalizable that target those main gap areas. Finally, to assist practitioners to apply the multi-level governance AI in organizations and the impact that it has on the industry level, we will translate into effective and responsible AI practices using a case study.

Index Terms—AI, AI ethics, trustworthy AI, AIMLOps, AIOps, software engineering, software architecture, pattern, best practice

I. INTRODUCTION

Aritificial Intelligence (AI) reshaped our lives, helped people make better predictions and take more informed and wise decisions. However, these high tech are still in there infancy, and there remains much promise for AI to promote innovation and address global challenges that people face.

Consecutively, ethical concerns and anxieties are fuelling around AI [1]. There are lots of enquiries on the trustworthiness and adoption of AI systems, including concerns about exacerbating inequality, digital divide, climate change and market concentration. Additionally, there are concerns that the use of AI may compromise human rights and values such as privacy. To address these concerns and ensure the responsible development and use of AI, a collaborative effort involving multiple stakeholders and international cooperation issued guidelines and ethical principles. Despite the creation of ethical guidelines for AI development inside organization, it can be challenging for developers to apply these principles in practical situations. These principles are often abstract and may not provide clear direction for specific implementation [2]. Therefore, more specific and actionable guidelines are needed to assist developers in implementing ethical considerations in their AI systems. It is important to bridge the gap between ethical principles and the algorithms used in AI systems to ensure responsible development. However, The architecture of an AI ecosystem consists of three layers: AI software supply chain, AI system, and operation infrastructure. It is challenging to show the contribution of each.

One work that was proposed is Responsible AI (RAI) Pattern Catalogue [3], which takes a pattern-oriented approach to promoting RAI in practice. Instead of solely focusing on ethical principles or AI algorithms, this catalogue focuses on design patterns that practitioners can apply to ensure that their AI systems are responsible throughout the software development process. The catalogue is organized into three categories: 1) governance patterns to establish multi-level governance, 2) process patterns to establish trustworthy development processes, and 3) product patterns to integrate responsible design into AI systems. In addition, it focuses on all aspect of the ecosystem (industry-level, organization-level and

team-level) without the planning of the design and the developement tools to support the navigation and utilisation of the RAI pattern catalogue.

In this paper, we take a different approach by focusing on the organization-level patterns at the system level rather than just the ethical principles or AI algorithms. This approach aims to integrate responsible design in organizations into final AI products by looking at the MLOps in a bigger picture and the design patterns that reshape the system. This is done with the intention of bridging the gap between the organizational-level and team-level and facilitating navigation using MLOps. We start off by looking at the main two levels of an organization with the addition to the team-level and examine the current available methods [6]–[10]. Then we make the links on where those methods meets and create the best practices using the multi-level governance patterns at the organization level. The overarching research question that has guided this study is:

> What is the multi-level governance pattern principle-to-practice proposed for responsible AI systems to bridge the gap between team-level and organization-level using MLOps?

The main contributions of this paper are as follows:

- Find the link between team-level governance patterns with the organization-level patterns.
- Suggest navigation and utilisation team-level governance patterns with the organization-level patterns.
- Explore a case study that suits this principle-to-practice multi-level governance pattern.

II. Related Work

The issue of creating AI that is ethically accountable has garnered a great deal of interest among both industrial and academic communities. To promote ethical AI practices, a multitude of AI ethics principles and guidelines numbering around 100 have been established by various entities including governments, companies, and organizations [12]. However, these guidelines are often too general and theoretical for individuals involved in the implementation of AI systems to apply in real-world scenarios.

There has been a concerted effort in the field of AI to address the challenges of RAI. One approach that has gained traction is the development of algorithm-level solutions. These solutions are designed to address specific aspects of the numerous high-level AI ethics principles and guidelines that have been established by various entities. By focusing on a subset of the principles, these algorithmic solutions aim to bring concrete and practical approaches to address some of the ethical concerns related to AI. One approach that developers used is by limiting user access and preventing reverse engineering or modifications to the system design. Rather than providing full access to

AI systems by running them locally, it is recommended to offer AI services through cloud-based platforms and manage interactions through APIs [11]. As an illustration, access to OpenAI's language model GPT-3 is limited to approved users who can only integrate it into their AI systems via API. Another example is Google Vision AI's facial recognition feature, which is limited to a select few celebrities and is only accessible through API. Despite these efforts, there have been instances where the algorithm has been exposed to the outside without proper internal review and verification, leading to potential issues with the responsible use of AI.

However, it's important to note that these algorithm-level solutions are just one part of the larger picture of RAI. Implementing them alone may not be enough to address all the ethical concerns related to AI, as the principles themselves are often complex and multifaceted. It requires a collaborative effort between researchers, developers, policymakers, and other stakeholders (board members, executives, managers) to ensure that AI is developed and used in an ethical and responsible manner.

III. Methodology

In order to build up the links of the multi-level governance for RAI systems within organizations, we first evaluated the available methods at the team and organizational level [3] to understand their strengths and limitations, and identified gaps that provided opportunities for improvement. As shown in Figure 1, the hierarchy of organization and team-level stakeholders in the industry is depicted on the left side of the illustration, providing a visual representation of the various levels of responsibility and decision making within the industry. The right side of the figure displays the current methods available, which are being utilized to support the operations and processes of the stakeholders.

The illustration provides a comprehensive overview of the stakeholders involved and the methods being utilized, offering insight into the strengths and limitations of the current methods. In addition, the use of XAI and RAI connectors, as shown in the illustration, can further optimize the operation of the current methods and support the efforts of the stakeholders. Utilizing these connectors can provide a more comprehensive and user-friendly experience, leading to improved outcomes and increased success for the organization and its teams.

Futhermore, we evaluated an examination of Machine Learning Operations (MLOps) technologies and tools for each stage of the project pipeline, as well as the roles involved [14]. In this examination, we identified the weakness of the method being used as the absence of XAI and RAI. The lack of XAI and RAI in the method being used can result in unintended consequences and decreased trust in the system. Therefore, it is important to consider incorporating these elements into any machine learning project to ensure accountability and transparency. To best
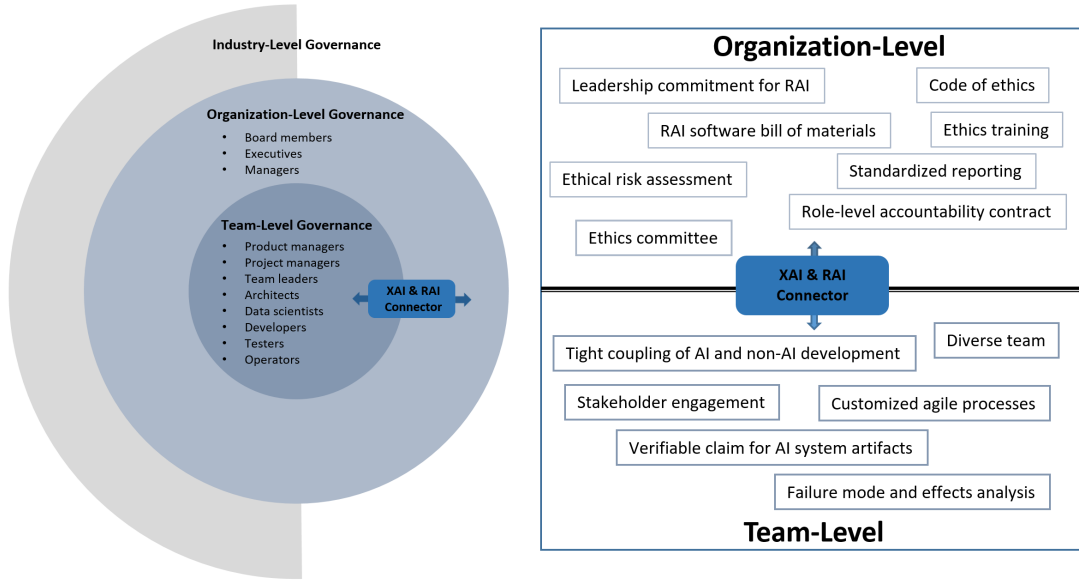
Fig. 1. A Deep Dive into the MLOps Workflow with the XRc Phase: Understanding Key Actor Roles and Responsibilities

to our knowledge there is no standard for implementing the multi-level governance pattern for RAI with XAI in MLOps.

XAI and RAI connector (XRc) can play a crucial role in connecting team-level governance to organization-level governance implementation in MLOps. By providing clear and understandable explanations for the decisions made by machine learning models, XAI helps to increase transparency and accountability at the team level. This can be especially important in complex projects involving multiple stakeholders and team members. RAI, on the other hand, helps to ensure that ethical and moral considerations are taken into account throughout the entire MLOps pipeline. This can involve creating policies and guidelines for RAI, as well as conducting risk assessments and impact evaluations. By incorporating RAI into MLOps, organizations can ensure that their use of AI aligns with their values and meets regulatory requirements.

By introducing XRc into the MLOps, organizations can bridge the gap between team-level governance and organization-level governance implementation in MLOps. This helps to ensure that AI systems are used in a responsible and ethical manner, while also providing a clear and transparent explanation of their decision-making process.

IV. BACKGROUND ON MLOPS WORKFLOW STAGES

Constructing a machine learning pipeline can be a challenging endeavor. The pipeline is often constructed incrementally with the assistance of tools that have limited integration capabilities. MLOps seeks to streamline this process by automating the pipeline. It serves as a combination of machine learning, data engineering, and DevOps practices, essentially streamlining and accelerating the operationalization of an ML model (including

building, testing, and releasing) by incorporating DevOps practices into the process. Determining which stage should be executed by which actor in the MLOps pipeline is not a straightforward task, and often requires multiple iterations to arrive at a suitable solution. However, through the examination of multiple studies, four major stages have been identified. Subsequently, we will outline each component in detail for that particular stage.

a) Data Management Phase: Maintaining the overall quality of project-specific data can be challenging due to restrictions posed by domain-specific limitations [53]. These limitations can affect the consistency of the relationships between attributes, the accuracy of historical records, and the reliability of state transitions [30]. To ensure that the data models are accurate and aligned with the goals and KPIs of the project, subject-specific experts such as domain experts play a crucial role . These experts provide the necessary problem-specific questions and objectives, as well as KPIs for the data models. Additionally, they are responsible for validating the potential data and machine learning models to ensure that they meet the requirements of the project.

In some organizations, Data Stewardship is a key entity responsible for overseeing data throughout the various stages of a project. With a focus on Data Quality Management and governance, various roles such as chief, business, and technical data stewards, as well as a data quality board, have been defined to guide the data governance framework of the organization [52]. The Data Stewardship plays an important role in maintaining resulting data, particularly when planning or wrapping up a project, often through a data management plan.

b) ML Preparation Phase: This set of functions in the ML preparation stage deals with classic ML preprocessing

tasks. Data quality is important and ensured by various roles with the help of data engineers and stewards. Implementing the ML model requires collaboration between data scientists, domain experts, and those responsible for defining the problem within the domain [42]. In summary, here are the functions in the ML preparation phase:

- ML Data Acquisition: The ML pipeline is fed with relevant data based on the prior declared data management plan and selection by the data engineer.
- Data Cleaning and Labeling: The input data is cleaned and labeled for ML operations with the help of data scientists and domain experts [30].
- Data Versioning: The separation of test, training, and validation data sets is crucial for the success of ML models and is achieved through data versioning.

c) ML Training Phase: The role of data scientists is crucial in the ML Pipeline Phase. They ensure the flexibility, scalability, and proper technology selection of the ML pipeline, while also working to improve model performance. They select the ML pipeline structure, algorithms, and hyperparameters through model versioning and validation, and are the main users in data processing for big data projects. AutoML techniques [44] and tools support data scientists and domain experts in efficiently selecting the ML pipeline and training the model. The process includes automation of feature preprocessing, model selection, and hyperparameter optimizationa. To sum up, the following functions are performed in the ML preparation phase:

- Pipeline Structure Search: The structure of an ML pipeline depends on the type of data (structured or unstructured) and the technique used to solve the problem (supervised, unsupervised, or semi-supervised learning). The specific performance metrics to be used must also be defined based on the specific domain-specific requirements of the problem being solved.
- Algorithm and Hyperparameter Selection: The choice of the most suitable ML algorithm for a problem is made by data scientists. The algorithm's performance can be improved by adjusting its hyperparameters, such as the number of layers in a neural network [36]. However, this process can be time-consuming and complex which addressed by AutoML in [44].
- Model Versioning: It is a way to keep track of the interdependencies between an ML model, its data, framework, and modeling procedure. It is important for reverting to previous models if there is a problem in production and for deploying the correct version at the right time. Model versioning increases accountability and is an essential component for managing complex ML models.

d) Deployment Phase: The deployment stage is a pivotal moment in the MLOps process. During this phase, software engineers are responsible for incorporating the approved models into the corresponding applications and ensuring the smooth operation of the entire system. To maintain this stability, MLOps engineers must continuously monitor the model, the application as a whole, and the data being used [42]. Another key player in this phase is the DevOps engineer, who is in charge of constructing, testing, and deploying the functioning system. In general, it is characterized by the following tasks:

- Integration of validated models into the relevant applications by software engineers.
- Maintenance of the operational stability of the entire system by MLOps engineers through continuous monitoring of the model, application, and data.
- Construction, testing, and deployment of the functioning system by DevOps engineers.

V. XAI and RAI Connector(XRc)

The integration of XRc into the MLOps pipeline may come with added overhead, however, it proves to be a valuable addition to the process as a whole. The addition of XRc not only reduces the risk of failure in RAI, but it also promotes efficiency by allowing for early detection of any problems with implementing organizational-level governance. This helps to avoid duplicating efforts and ensures that the RAI is being developed effectively. As shown in Figure 2, XRc has been inserted between the ML pipeline phase and the ML deployment phase in order to analyze the changes before migrating into the application incorporating the organization level governance into the process. Let's now delve into the sub-phases of XRc.

A. Model Type Identification

Both dynamic and static identification methods can be used to identify the type of machine learning model, with dynamic methods involving examination of the model's API or performance, and static methods involving examination of the code and architecture used to implement the model.

a) Static Identification Method: In code, the type of machine learning model can be identified by examining the architecture, algorithms, and libraries used to implement the model. Understanding the architecture of the model, such as the number of hidden layers or the presence of decision trees, can give a good indication of the type of machine learning model used. Additionally, many machine learning libraries provide pre-built models with clear documentation that specify the type of model being used. The documentation for these libraries usually clearly states the type of model being used. For example, in the scikit-learn library [37], the use of 'LogisticRegression' class for logistic regression, which is a supervised learning algorithm, or the 'KMeans' class for k-means clustering, which is an unsupervised learning algorithm.

b) Dynamic Identification Method: There are two main ways to dynamically detect the type of machine learning model:
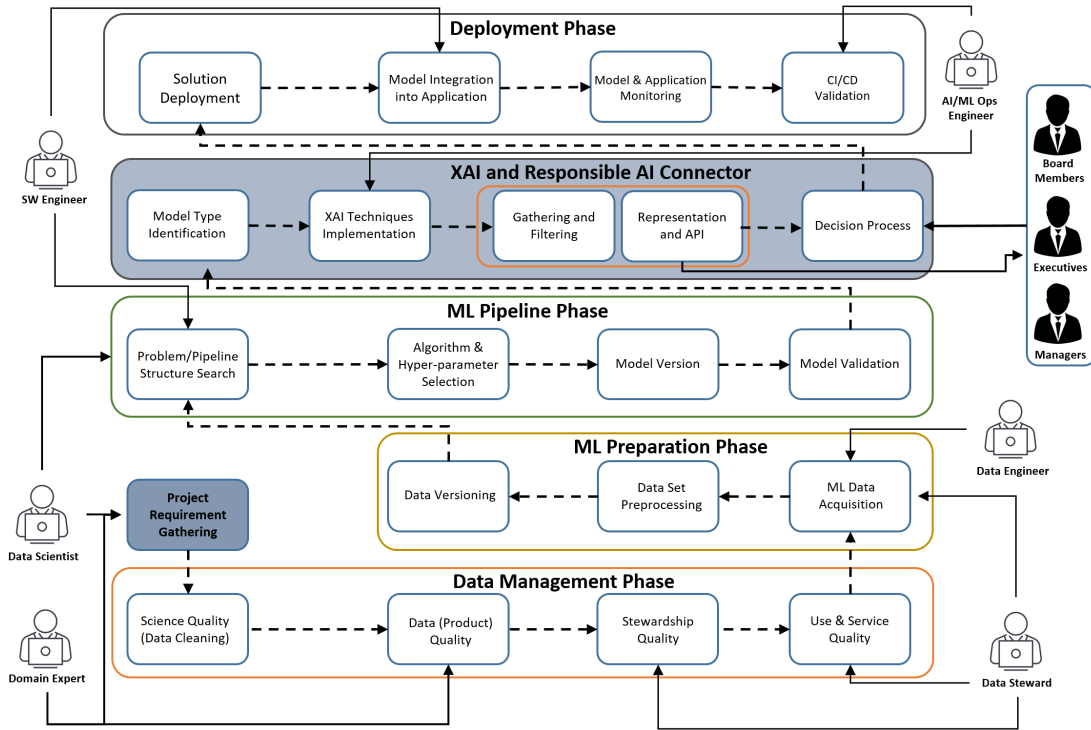
Fig. 2. The XRc to bridge explainable and responsible operations with machine learning lifecycle

- Examining the model's API or functions: This involves looking at the functions that the model exposes, such as the 'predict' function, and determining the type of model based on the inputs and outputs of the function.
- Examining the model's performance: This involves evaluating the model's performance on a known dataset and determining the type of model based on the performance metrics and results obtained.

The two dynamic methods can be useful in different scenarios, and the choice of method will depend on the specific requirements of the use case. For example, if the access to the model's API or functions is given, examining the API or functions might be the easiest and most straightforward way to determine the type of model. If there is no access to the API or functions, evaluating the model's performance on a known dataset might be the only option available.

B. XAI techniques Implementation

XAI techniques is implemented by AI/MLOps engineers taking into consideration the ML algorithm and the different audiences. It may require a trade-off between technical detail and simplicity, transparency and fairness, and other factors.

a) Audience Evaluation: While the audiences is mainly at the organization level, AI/MLOps Engineers can categories their audience into two:

- End Users: It may require simple and understandable explanations of the decisions made by a machine learning model. XAI techniques designed for this audience might include rule-based systems, decision trees, or prototype-based explanations that provide examples that are representative of the decisions made by a machine learning model.
- Regulators: It may require explanations of the decisions made by a machine learning model to ensure that the model is fair, transparent, and in compliance with regulations and ethical standards. XAI techniques designed for this audience might include model-agnostic techniques, such as Local Interpretable Model-Agnostic Explanation (LIME) [54] or SHapley Additive exPlanation (SHAP) [55], that provide explanations for the predictions made by any machine learning model.

b) ML Algorithm Evaluation: It is not necessary for an AI/MLOps engineer to have a deep understanding of the specific ML algorithm in order to choose an XAI method. However, having a general understanding of the ML algorithm and its strengths and weaknesses can be helpful in choosing an appropriate XAI method. In addition, since some XAI methods can be used and integrated together, AI/MLOps engineer have to consider the broad range of XAI techniques. Those methods can be categorized into the following categories [13]: Model-Agnostic Techniques: Model-agnostic XAI techniques are techniques that can be applied to any machine learning

model, regardless of the underlying algorithm or architecture. Examples of model-agnostic XAI techniques include LIME and SHAP.

- Model-Specific Techniques: Model-specific XAI techniques [13] are techniques that are designed specifically for a particular type of machine learning model, such as decision trees or neural networks. Examples of model-specific XAI techniques include saliency maps for neural networks and decision trees for decision trees.
- Post-Hoc Techniques: Post-hoc XAI techniques [13] are techniques that are applied to a trained machine learning model after it has been trained, in order to provide explanations for its decisions and actions. Examples of post-hoc XAI techniques include LIME, SHAP, and saliency maps [47].
- Integrative Techniques: Integrative XAI techniques are techniques that involve integrating XAI into the training process of a machine learning model, so that explanations can be generated as part of the model's normal operation.

These categories are not mutually exclusive and some XAI techniques may fall into multiple categories. The choice of XAI technique will depend on the specific requirements of the use case and the type of model being used.

C. Gathering and Filtering

Gathering and filtering XAI methods involves selecting a subset of XAI methods that are appropriate for a specific use case, and then integrating and combining those methods to provide a more comprehensive and effective explanation of the decisions made by a machine learning model. To effectively utilize XAI methods for a specific use case, here are the steps to gather and filter XAI methods:

- Define the requirements: AI/MLOps engineers should start by defining the requirements based on the audience they are trying to target and the type of model being used. This will help to determine which XAI methods are most appropriate for the use case.
- Gather XAI methods: Next, gathering a set of XAI methods that are appropriate and generated by the XAI technique Implementation. This may require back and forth to add, remove certain XAI techniques.
- Filter XAI methods: Once a set of XAI methods is gathered, filter the methods based on the specific requirements to the use case. Consider factors such as the complexity of the method, the amount of computational resources required, and the amount of technical detail that is appropriate for the audience.
- Combine XAI methods: After filtering the XAI methods, MLOps Engineers might consider combining methods to provide a more comprehensive and effective explanation of the decisions made by the machine learning model. They can use techniques such as

multi-modal explanations, ensemble explanations, or hybrid explanations to combine XAI methods.

D. Representation and API

XAI representation refers to the format in which the explanations generated by XAI techniques are presented to the user. The representation can be in the form of text, visualizations, or other forms of data, and it depends on the specific requirements of the use case and the target audience.

An XAI API, or Application Programming Interface, is a set of protocols and tools for building software applications that provide access to XAI capabilities. An XAI API allows developers to easily integrate XAI techniques into their applications and provide explanations for the decisions made by machine learning models. For example, an XAI API might provide a set of functions that allow developers to generate explanations for specific predictions made by a machine learning model, or to visualize the model's decision-making process. The XAI API might also provide a set of data structures and protocols for representing the explanations generated by XAI techniques, such as text, visualizations, or other forms of data. Here are some ways to make the XAI API interactive:

- Visualizations: Use interactive visualizations, such as heatmaps, bar charts, and scatter plots, to help stakeholders understand the explanations generated by the XAI API. This can include interactive visualizations of the model's behavior, such as feature importance, or visualizations of individual predictions, such as decision trees.
- User Input: Allow stakeholders to provide input to the XAI API, such as selecting specific predictions to explain, or adjusting parameters of the explanations. This can help stakeholders to better understand the explanations generated by the API, and to tailor the explanations to their specific needs.
- Dynamic Explanations: Provide dynamic explanations that change based on user input or other factors, such as the type of model being used, or the specific prediction being explained. This can help stakeholders to better understand the explanations generated by the API, and to see how changes in the model or input data affect the explanations.
- Feedback Mechanisms: Provide feedback mechanisms that allow stakeholders to provide feedback on the explanations generated by the XAI API. This can include simple feedback forms, or more complex mechanisms, such as a feedback rating system. This can help organizations to improve the quality of the explanations generated by the API, and to better meet the needs of stakeholders.

In a nutshell, XAI representation refers to the format in which XAI explanations are presented to the user, while an XAI API provides a set of protocols and tools for integrating XAI techniques into software applications and

TABLE I
Mapping of XAI Service Architecture Design to XRc MLOps Phase

XAI Microservices & Provenance Ref	Components of XAI Microservices	XRc Sub-phase Mapping
User Interface	RESTful API Web Portal Command Line	Representation and API
Coordination Center	Microservice Management XAI Operation Management	Gathering and Filtering
Microservice Layer	Data Representation Data Processing AI Model XAI Method Evaluation Metric	XAI Techniques Implementation
Data Persistent Layer	Model Explanation Result XAI Evaluation Result	XAI Techniques Implementation
	Machine Learning Data Provenance Data	Model Type Identification

providing explanations for the decisions made by machine learning models. The choice of XAI representation and API will depend on the specific requirements of the use case and the target audience.

E. Decision Process

Incorporating an XAI API into the decision-making process at the organizational level can help stakeholders make informed decisions about whether to proceed with or stop a ML algorithm based on the transparency and fairness of the model. To make the XAI API interactive, SW Engineers and MLOps can add features that allow stakeholders to interact with the explanations generated by the API.

VI. Case Study: Flexible XAI Service for Cloud AI Feature Contributions

Since the XRc invoked in MLOps can target many different MLOps pipelines and cannot be easily specified, it is essential to present a specific case study to demonstrate the usability of these techniques in MLOps. In this section, we will elaborate on the Flexible XAI Service for Cloud AI Feature case study that will pick one MLOps pipeline and show the effectiveness of XRc techniques in improving its transparency and fairness.

It is important to note that this case study will focus solely on the implementation, gathering and filtering, and representation and API [11] aspects of XRc phase in MLOps pipeline. While there are other components within the XRc phase, these sub-phases are considered to be the most critical elements within the XRc phase.

A. Mapping XAI Service Architecture Design and XRc MLOps phase

As depicted, the development of XAI operations as microservices resulted in a reference architecture composed of four layers: the user interface layer, the coordination center layer, the core microservice layer, and the data persistence layer. This reference architecture can be mapped to the XRc phase, as shown in Table I.

a) User Interface: The XAI service architecture has a user interface that allows users to set up XAI tasks or connect them to pipelines. The interface offers three forms of access: a web portal, RESTful APIs, and a command line. The web portal includes features such as task configuration, explanation summaries, and data visualization. RESTful APIs support developers in building new XAI services and executing pipelines. The command line interface allows direct communication with the backend server. These features support the Representation and API aspect of MLOps.

b) Coordination Center: The coordination center in the XAI service architecture plays a crucial role in Gathering and Filtering in XRc Phase in the MLOps. The microservice management function is responsible for managing the life cycle of each microservice and documenting their endpoints and functional types. The XAI operation management includes a task executor that executes XAI tasks based on task sheets that contain information about the involved services, data set, and task parameters [18]. This information is gathered and filtered by the coordination center to ensure that the correct microservices are utilized for each task. The data representation component retrieves information from the data persistence layer and presents it to the user, further supporting the Gathering and Filtering aspect of XRc Phase in MLOps.

c) Microservice Layer: The encapsulation of diverse data sets, AI models, XAI methods, and evaluation metrics into microservices as described in the previous paragraph is a key aspect of the XAI Techniques Implementation subphase of the XRc phase in MLOps. The use of microservices simplifies the integration of these entities and supports the efficient execution of XAI tasks within the MLOps pipeline. The Data Processing, AI Model, XAI
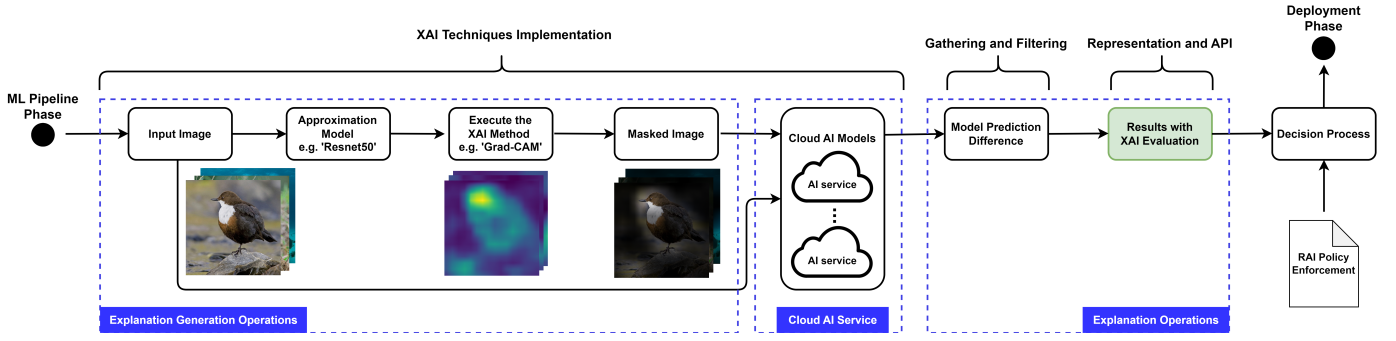
Fig. 3. Mapping the Illustrated Integration of Cutting-Edge Cloud AI Services, State-of-the-Art Deep Learning Models, and Post Hoc Explainability Methods in the XRc Phase of MLOps.

Method, and Evaluation microservices each play a specific role in supporting the implementation of XAI techniques in the XRc phase, making them crucial components of the XAI Techniques Implementation subphase of the overall MLOps process.

d) Data Persistent Layer: The data persistence layer in the XAI service is responsible for managing the storage and retrieval of data sets, XAI operation data, explanation summaries, and evaluation results. This layer plays a crucial role in the XRc phase of MLOps by providing a unified system for storing and accessing the necessary information for XAI operations. The design of XAI provenance metadata is also introduced in a later section, further enhancing the efficiency and organization of the XAI process. The data persistence layer is a key component of both the XAI Techniques Implementation and Model Type Identification subphases of the XRc phase in MLOps. The ability to easily add new AI models, data sets, or XAI methods further underscores the importance of the data persistence layer in supporting the overall MLOps process. In the next section, we will delve deeper into the implementation of the Post Hoc XAI with feature contribution explanation, highlighting its practical applications and benefits in the XRc Phase of MLOps.

B. Post Hoc XAI Methods with Feature Contribution Explanation

Post-hoc XAI methods for feature influence and causality can be applied to computer vision AI services as they provide a way to understand how the deep learning models make their predictions. These methods can be used to approximate the internal workings of the AI models and provide insights into the contribution of each feature to the prediction result. This information can be useful for improving the accuracy of the models, as well as for building trust with users and stakeholders who may be concerned about the transparency of AI systems. Additionally, the availability of state-of-the-art deep learning models and XAI methods for computer vision in the literature makes it easier to adopt these methods for computer vision AI services.

As depicted in Figure 3, the Explanation Generation Operation and the Cloud AI Service are a part of the XAI Techniques Implementation. This makes sense since the process involves inputting images into a cloud-based AI service in the first branch, which outputs the prediction results. In the second branch, the images are processed by an approximation model such as ResNet or DenseNet [50] , [49]. XAI methods specific to the model [4] are then utilized to generate a salience map that highlights the most significant features of the image. This map is applied to the original image, creating augmented images that are then input into the cloud-based AI service. As a result, the prediction results from the cloud-based AI service on these augmented images will differ from the original prediction due to the changes in the features.

Since the XAI techniques implementation has taken place, it is possible to do more gathering and filtering by using the computed values. This done through additional analysis of the impact of feature changes on the outcome. As such, A clear explanation can be generated and its reliability and coherence can be assessed based on prior research on explanation evaluation. The closer the predictions generated by the two models, i.e., the cloud-based AI service and the approximation models such as ResNet and DenseNet the more similarities there are in the features that influence their predictions.

Finally, given that RESTful APIs provide the sole means of accessing most services, the combination of relevant components and processes shown in Figure 3 illustrate the type of services utilizing APIs specified on service endpoints. This serves as a significant factor in advocating for the implementation of microservices for XAI operations. This cleary a direct mapping to the XRc representation and API phase.

C. Reflection

As shown in Figure 3, integrating an XRc phase between the ML Pipeline and Deployment phases is a good practice in order to ensure that AI models are explainable, transparent, and accountable. This XRc phase can involve various steps such as auditing the dataset, validating the

model, analyzing the fairness and bias, and generating human-interpretable explanations for the model's predictions. In addition, integrating XRc decision processes into existing tools and workflows can be done in several ways, depending on the specific use case and the tools and technologies involved. Here are some possible approaches:

- API integration: An API can be used to automatically generate human-interpretable explanations of the model's predictions, which can then be integrated into existing applications or dashboards. This can involve setting up an API endpoint that receives input data, passes it through the ML model, and returns the predictions along with explanations for each prediction.
- Email with survey response: Integrating XRc decision processes into email with survey response can involve setting up an automated email workflow that sends out surveys to stakeholders and collects feedback on the model's predictions. The survey responses can be analyzed and used to improve the model's performance, and the updated model can be integrated back into the decision-making process.
- Knowledge base: Incorporating a knowledge base into the XRc decision process can help to provide stakeholders with a centralized source of information that they can reference at any time. This can be particularly useful for stakeholders who may not have a deep understanding of machine learning or the specific model being used.

In each of these approaches, it's important to ensure that the XRc decision process is integrated seamlessly into the existing tools and workflows, and that the output is presented in a way that is easily understandable and actionable for stakeholders. It's also important to ensure that the XRc decision process is audited regularly to ensure that it continues to meet the organization's ethical and responsible standards.

VII. Conclusion

The connection between team-level governance patterns and organization-level patterns can be effectively established through the use of MLOps. The XRc phase provides a clear connection between the two levels of governance, allowing for the navigation and utilization of both patterns. The Design Explanation Microservices and Provenance case study serves as a practical example of this principle-to-practice multi-level governance pattern, showcasing how XAI techniques can be effectively implemented in real-world scenarios to support both team and organizational governance goals. By utilizing the XRc phase in MLOps and examining the results of the case study, organizations can gain a better understanding of how to bridge the gap between team-level and organizational-level governance in a way that is both effective and efficient.

Future work in this area could involve exploring new and innovative XAI techniques and methods to further en-

hance the efficiency and effectiveness of the XRc phase in MLOps. Additionally, continued research into the relationship between team-level and organization-level governance patterns in the context of MLOps will be important in order to fully understand the best practices and strategies for bridging the gap between these two levels of governance.

References

[1] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *CoRR*, vol. abs/1908.09635, 2019. [Online]. Available: http://arxiv.org/abs/1908.09635

[2] Q. Lu, L. Zhu, X. Xu, J. Whittle, D. Douglas, and C. Sanderson, "Software engineering for responsible AI: an empirical study and operationalised patterns," *CoRR*, vol. abs/2111.09478, 2021.

[3] Q. Lu, L. Zhu, X. Xu, J. Whittle, D. Zowghi, and A. Jacquet, "Responsible ai pattern catalogue: A multivocal literature review," 2022.

[4] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," *CoRR*, vol. abs/1610.02391, 2016. [Online]. Available: http://arxiv.org/abs/1610.02391

[5] F. Hussain, R. Hussain, B. Noye, and S. Sharieh, "Enterprise API security and GDPR compliance: Design and implementation perspective," *CoRR*, vol. abs/1909.08048, 2019. [Online]. Available: http://arxiv.org/abs/1909.08048

[6] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, oct 2020.

[7] ——, "Responsible ai: Bridging from ethics to practice," *Commun. ACM*, vol. 64, no. 8, p. 32–35, jul 2021.

[8] J. S. Borg, "Four investment areas for ethical ai: Transdisciplinary opportunities to close the publication-to-practice gap," *Big Data & Society*, vol. 8, no. 2, p. 20539517211040197, 2021.

[9] W. Hussain, M. Shahin, R. Hoda, J. Whittle, H. Perera, A. Nurwidyantoro, R. Shams, and G. Oliver, "How can human values be addressed in agile methods? a case study on safe," 02 2021.

[10] Q. Lu, L. Zhu, X. Xu, J. Whittle, and Z. Xing, "Towards a roadmap on software engineering for responsible ai," 03 2022.

[11] T. Shevlane, "Structured access to AI capabilities: an emerging paradigm for safe AI deployment," *CoRR*, vol. abs/2201.05159, 2022.

[12] A. Jobin, M. Ienca, and E. Vayena, "Artificial intelligence: the global landscape of ethics guidelines," *CoRR*, vol. abs/1906.11668, 2019. [Online]. Available: http://arxiv.org/abs/1906.11668

[13] Q. Ai and L. N. Ramasamy, "Model-agnostic vs. model-intrinsic interpretability for explainable product search," *CoRR*, vol. abs/2108.05317, 2021.

[14] P. Ruf, M. Madan, C. Reich, and D. Ould-Abdeslam, "Demystifying mlops and presenting a recipe for the selection of open-source tools," *Applied Sciences*, vol. 11, no. 19, 2021.

[15] S. Shrivastava, D. Patel, N. Zhou, A. Iyengar, and A. Bhamidipaty, "Dqlearn : A toolkit for structured data quality learning," in *2020 IEEE International Conference on Big Data (IEEE BigData 2020), Atlanta, GA, USA, December 10-13, 2020*, X. Wu, C. Jermaine, L. Xiong, X. Hu, O. Kotevska, S. Lu, W. Xu, S. Aluru, C. Zhai, E. Al-Masri, Z. Chen, and J. Saltz, Eds. IEEE, 2020, pp. 1644–1653.

[16] E. Raj, *Engineering MLOps: Rapidly build, test, and manage production-ready machine learning life cycles at scale*. Packt Publishing, 2021.

[17] I. Karamitsos, S. Albarhami, and C. Apostolopoulos, "Applying devops practices of continuous automation for machine learning," *Information*, vol. 11, no. 7, 2020.

[18] S. Mäkinen, H. Skogström, E. Laaksonen, and T. Mikkonen, "Who needs mlops: What data scientists seek to accomplish and how can mlops help?" 2021.

[19] "Mlops workload orchestrator," accessed on August 12, 2022.

[20] "Aws mlops framework," accessed on August 13, 2022.

[21] S. Sharma, *The DevOps Adoption Playbook: A Guide to Adopting DevOps in a Multi-Speed IT Enterprise*. Wiley, 2017.

[22] V. Gudivada, A. Apon, and J. Ding, "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations," *International Journal on Advances in Software*, vol. 10, pp. 1–20, 07 2017.

[23] P. Mattson, C. Cheng, C. Coleman, G. Diamos, P. Micikevicius, D. A. Patterson, H. Tang, G. Wei, P. Bailis, V. Bittorf, D. Brooks, D. Chen, D. Dutta, U. Gupta, K. M. Hazelwood, A. Hock, X. Huang, B. Jia, D. Kang, D. Kanter, N. Kumar, J. Liao, G. Ma, D. Narayanan, T. Oguntebi, G. Pekhimenko, L. Pentecost, V. J. Reddi, T. Robie, T. S. John, C. Wu, L. Xu, C. Young, and M. Zaharia, "Mlperf training benchmark," *CoRR*, vol. abs/1910.01500, 2019. [Online]. Available: http://arxiv.org/abs/1910.01500

[24] G. Fursin, H. Guillou, and N. Essayan, "Codereef: an open platform for portable mlops, reusable automation actions and reproducible benchmarking," *CoRR*, vol. abs/2001.07935, 2020.

[25] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.

[26] J. Zhao, M. Mathieu, R. Goroshin, and Y. LeCun, "Stacked what-where auto-encoders," 2015.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[28] E. R. Sparks, A. Talwalkar, V. Smith, J. Kottalam, X. Pan, J. E. Gonzalez, M. J. Franklin, M. I. Jordan, and T. Kraska, "MLI: an API for distributed machine learning," *CoRR*, vol. abs/1310.5426, 2013.

[29] M. A. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, J. M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Fast and interactive analytics over hadoop data with spark," *login Usenix Mag.*, vol. 37, 2012.

[30] I. Taleb, M. A. Serhani, and R. Dssouli, "Big data quality: A survey," in *2018 IEEE International Congress on Big Data (BigData Congress)*. IEEE, 2018, pp. 166–173.

[31] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2438–2455, 2021.

[32] L. N. Tidjon and F. Khomh, "Threat assessment in machine learning based systems," 2022.

[33] P. Fukas, J. Rebstadt, F. Remark, and O. Thomas, "Developing an artificial intelligence maturity model for auditing," 06 2021.

[34] R. Sadiq, N. Mohd Satar, A. H. Abd Rahman, and S. Goudarzi, "Artificial intelligence maturity model: A systematic literature review," *PeerJ Computer Science*, vol. 7, 08 2021.

[35] B. Fish and L. Stark, "Reflexive design for fairness and other human values in formal models," 07 2021, pp. 89–99.

[36] D. Kißkalt, A. Mayr, B. Lutz, A. Rögele, and J. Franke, "Streamlining the development of data-driven industrial applications by automated machine learning," *Procedia CIRP*, vol. 93, pp. 401–406, 2020.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[38] I. Naja, M. Markovic, P. Edwards, and C. Cottrill, "Correction to: A semantic framework to support ai system accountability and audit," in *The Semantic Web*, R. Verborgh, K. Hose, H. Paulheim, P.-A. Champin, M. Maleshkova, O. Corcho, P. Ristoski, and M. Alam, Eds. Cham: Springer International Publishing, 2021, pp. C1–C1.

[39] S. Jentzsch, S. Höhn, and N. Hochgeschwender, *Conversational Interfaces for Explainable AI: A Human-Centred Approach*, 09 2019, pp. 77–92.

[40] A. Perera, A. Aleti, C. Tantithamthavorn, J. Jiarpakdee, B. Turhan, L. Kuhn, and K. Walker, "Search-based fairness testing for regression-based machine learning systems," *Empirical Software Engineering*, vol. 27, 05 2022.

[41] I. Guzey, O. Ucar, N. A. Ciftdemir, and B. Acunas, "Context-dependent explainability and contestability for trustworthy medical artificial intelligence: Misclassification identification of morbidity recognition models in preterm infants," 2022.

[42] M. Treveil, N. Omont, C. Stenac, K. Lefevre, D. Phan, J. Zentici, A. Lavoillotte, M. Miyazaki, and L. Heidmann, *Introducing MLOps*. O'Reilly Media, 2020.

[43] H. Mehmood, P. Kostakos, M. Cortes, T. Anagnostopoulos, S. Pirttikangas, and E. Gilman, "Concept drift adaptation techniques in distributed environment for real-world data streams," *Smart Cities*, vol. 4, no. 1, pp. 349–371, 2021.

[44] F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., *Automatic Machine Learning: Methods, Systems, Challenges*. Springer, 2019.

[45] M. Feurer, A. Klein, K. Eggensperger, J. T. Springenberg, M. Blum, and F. Hutter, "Auto-sklearn: Efficient and robust automated machine learning," F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Springer, 2019, pp. 123–143.

[46] N. Muralidhar, S. Muthiah, P. Butler, M. Jain, Y. Yu, K. Burne, W. Li, D. Jones, P. Arunachalam, H. S. McCormick *et al.*, "Using antipatterns to avoid mlops mistakes," *arXiv preprint arXiv:2107.00079*, 2021.

[47] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[48] T. Granlund, A. Kopponen, V. Stirbu, L. Myllyaho, and T. Mikkonen, "Mlops challenges in multi-organization setup: Experiences from two real-world cases," in *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*. IEEE, 2021, pp. 82–88.

[49] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[51] D. A. Tamburri, "Sustainable mlops: Trends and challenges," in *2020 22nd international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)*. IEEE, 2020, pp. 17–23.

[52] B. Mons, *Data stewardship for open science: Implementing FAIR principles*. Chapman and Hall/CRC, 2018.

[53] A. Maydanchik, *Data quality assessment*. Technics publications, 2007.

[54] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[55] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017. [Online]. Available: https://arxiv.org/abs/1705.07874