Conference Paper Title*

Elie Neghawi
Electrical and Computer Engineering
Concordia University
Montreal, Canada
e_negh@encs.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 (MLG) 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 MLG 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 two case studies.**

***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 Pattern Catalogue [3], which takes a pattern-oriented approach to promoting responsible AI 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 (MLG), 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 Responsible AI 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 bigger picture and the design patterns that shape the system as a whole. This is done with the intention of bridging the gap between the organizational-level and team-level and facilitating navigation. 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 [5]–[9]. Then we make the links on where those methods meets and create the best practices using the MLG patterns at the organization level. The overarching research question that has guided this study is:

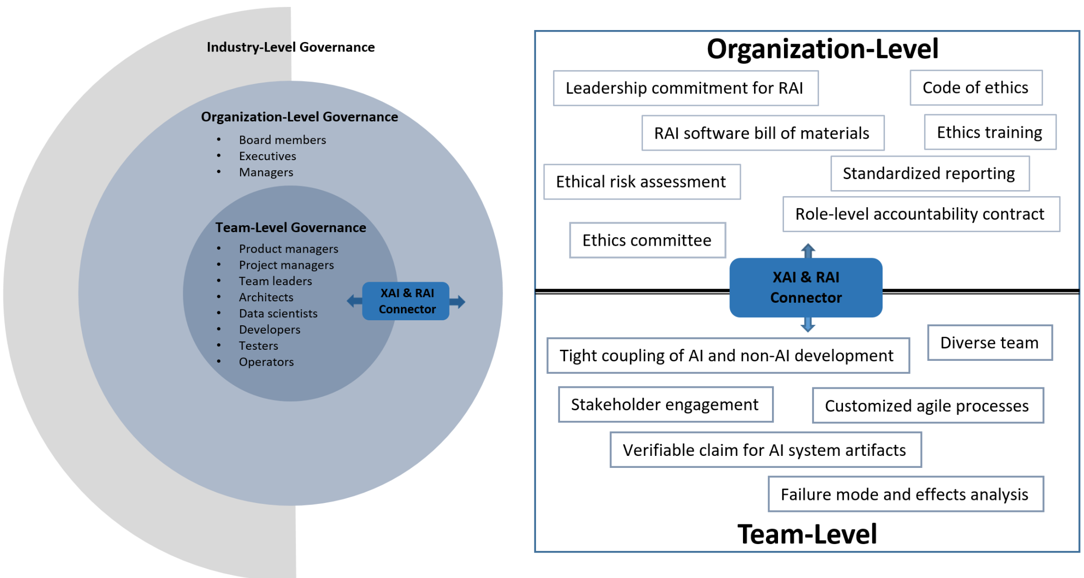

Fig. 1. Transition from traditional to the current approach

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?

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 two 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 [11]. 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 responsible AI. 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 [10]. 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 responsible AI. 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 responsible AI systems within organizations, we first evaluated the available methods at the team and organizational level 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 [12]. 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 to our knowledge there is no standard for implementing the MLG 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 responsible AI, 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. These limitations can affect the consistency of the relationships between attributes, the accuracy of historical records, and the reliability of state transitions. 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. 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 (DMP).

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. 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.
- 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,

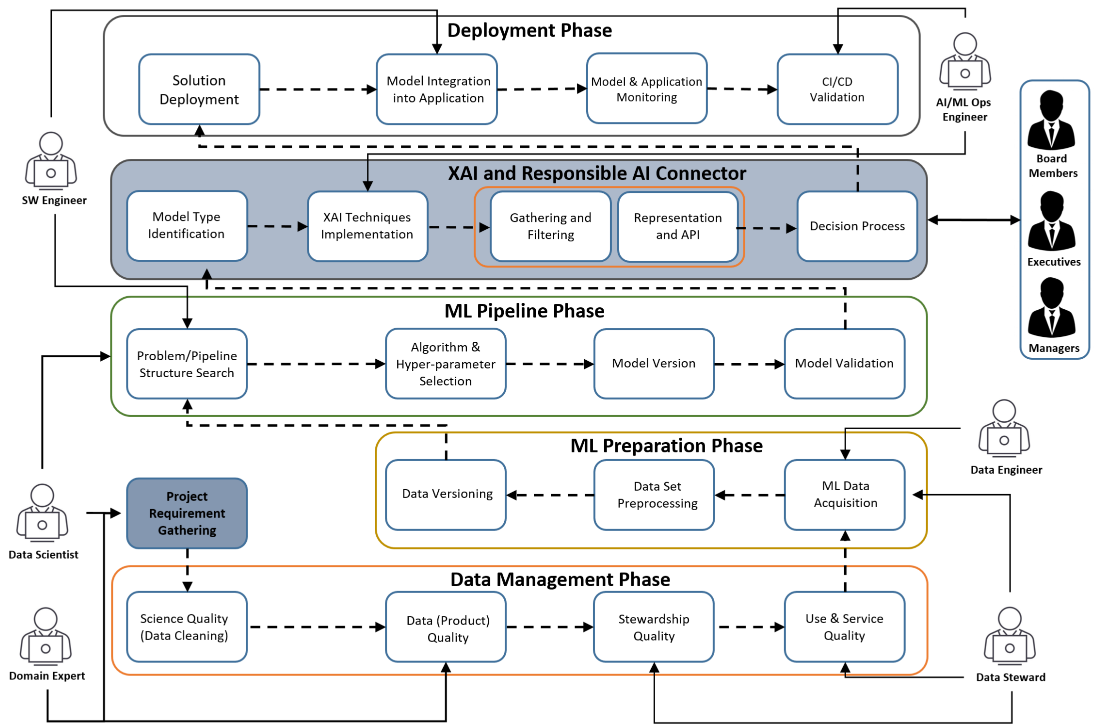

Fig. 2. Transition from traditional to the current approach

and hyperparameters through model versioning and validation, and are the main users in data processing for big data projects. AutoML techniques 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. However, this process can be time-consuming and complex which addressed by AutoML.
- 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. 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 responsible AI, 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 responsible AI 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 subphases of XRc.

A. Model Type Identification

Both dynamic and static 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 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 you 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, you can use the '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 Method: There are two main ways to dynamically detect the type of machine learning model:

- 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 your use case. For example, if you have access to the model's API or functions, examining the API or functions might be the easiest and most straightforward way to determine the type of model. If you don't have 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

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. **??**", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number 'citations consecutively within brackets . The sentence punctuation follows the bracket . Refer simply to the reference number, as in —do not use "Ref. " or "reference " except at the beginning of a sentence: "Reference was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" . Papers that have been accepted for publication should be cited as "in press" . Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation .

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. [Online]. Available: https://arxiv.org/abs/2111.09478

[3] Q. Lu, L. Zhu, X. Xu, J. Whittle, D. Zowghi, and A. Jacquet, "Responsible ai pattern catalogue: A multivocal literature review," 2022. [Online]. Available: https://arxiv.org/abs/2209.04963

[4] 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

[5] 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. [Online]. Available: https://doi.org/10.1145/3419764

[6] ——, "Responsible ai: Bridging from ethics to practice," *Commun. ACM*, vol. 64, no. 8, p. 32–35, jul 2021. [Online]. Available: https://doi.org/10.1145/3445973

[7] 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. [Online]. Available: https://doi.org/10.1177/20539517211040197

[8] 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.

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

[10] T. Shevlane, "Structured access to AI capabilities: an emerging paradigm for safe AI deployment," *CoRR*, vol. abs/2201.05159, 2022. [Online]. Available: https://arxiv.org/abs/2201.05159

[11] 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

[12] 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. [Online]. Available: https://www.mdpi.com/2076-3417/11/19/8861