Mapping and Mitigating the Pains of Responsible Artificial Intelligence Implementation: An Analysis of Swiss Organizations

Eleonora Viganò
Digital Society Initiative
University of Zurich,
Zurich, Switzerland
Swiss Institute for Entrepreneurship
University of Applied Sciences of the Grisons
Chur, Switzerland
eleonora.vigano@uzh.ch

Christian Hauser
Swiss Institute for Entrepreneurship
University of Applied Sciences of the Grisons
Chur, Switzerland
christian.hauser@fhgr.ch

Eugenia Cacciatori
Bayes Business School of City
University of London
London, United Kingdom
Eugenia.Cacciatori.1@city.ac.uk

Andrea Ferrario*
Institute of Biomedical Ethics and History of Medicine
University of Zurich,
ETH Zurich
Zurich, Switzerland
aferrario@ethz.ch

Jana Sedlakova*
Department of Informatics,
University of Zurich,
Digital Society Initiative
University of Zurich,
Zurich, Switzerland
sedlakova@ifi.uzh.ch

* The fourth and fifth authors contributed equally to the paper

***Abstract*—Despite the proliferation of Responsible Artificial Intelligence (RAI) principles, organizations struggle to translate them into practical implementation. This study investigates the challenges Swiss organizations face in implementing RAI through qualitative interviews with industry professionals and academic experts, complemented by a multi-stakeholder workshop. We first identify five critical *pain points* hampering RAI implementation: economic constraints, structural and procedural barriers, conceptual and technical challenges, cultural and behavioral resistance, and regulatory uncertainty. Then we propose the *Control-Tangibility Framework*, a novel framework that maps pain points along two fundamental dimensions: organizational control and challenge tangibility. Our framework provides organizations with a structured methodology to prioritize RAI efforts by considering both their ability to influence change and their capacity to observe aspects of the challenges. Furthermore, we provide practical insights for developing targeted implementation strategies that bridge the gap between ethical principles and operational practices. Our findings suggest that successful RAI implementation requires moving beyond compliance-focused approaches toward a comprehensive organizational transformation, supported by systematic assessment and prioritization of implementation challenges.**

Keywords— Responsible AI, Implementation Challenges, AI Governance, Trustworthy AI, Ethical AI

I. Introduction

Artificial Intelligence (AI) has reached a development point where rapid technological progress is accompanied by increasing adoption of AI across multiple industries. As AI systems have become prevalent in society and started to substantially and tangibly affect people's everyday lives, concerns about the ethical implications of their use are also growing. As a result, on the one hand, governments are working on regulatory frameworks to avoid or minimize the negative impact of AI on individuals, society, and the environment (e.g., the EU AI Act). On the other hand, in recent years, international organizations, private companies, and academic scholars published ethical frameworks and guidelines to regulate the development, deployment, and use of AI (e.g., [1], [2], [3], [4]). In this context, the term "Responsible AI" (RAI) is used to refer to an approach that prioritizes ethical considerations in the design, development, and deployment of AI technology in our societies [6], [7].[1] RAI aims to ensure AI creates positive societal impact while upholding principles such as fairness, transparency, and accountability.

The ethical frameworks and guidelines for AI have been criticized for providing only high-level norms without practical guidance for implementation [8], [9], [10], [11], [12]. Moreover, currently there exists no clear sector-specific approach to RAI in, for instance, financial services, healthcare, or manufacturing. In response, several businesses and scholars have proposed specific procedures and tools for applying RAI effectively (for a review, see [13]).

Notwithstanding the proliferation of ethical frameworks and the attempts to operationalize them, companies still struggle to integrate RAI into their everyday workflow [11], [14], [15], [16], [17]. Further, ethics-driven practical approaches to deliver AI-infused services and products, such as value sensitive design [18], lack contextualization and are not immediately applicable to industry. Moreover, the systematic analysis of the RAI implementation challenges received limited attention in the literature. As a result, organizations risk designing, developing, and using AI

[1]RAI is sometimes referred to as "ethical AI" or "trustworthy AI." The term "responsible" emphasizes the need for finding practical ways to address ethical challenges stemming from AI [5]. For an example of RAI policy, the interested reader may consult the *Montréal Declaration for a Responsible Development of Artificial Intelligence*, available at https://montrealdeclaration-responsibleai.com/the-declaration/.

This work was supported by the *Innovation Booster Databooster*, the Ethics Community of the Digital Society Initiative, and the companies participating in the Innovation Booster Databooster.

systems that can harm individuals, society, and the environment. Harmful AI systems can in turn create significant strategic, operational, and reputational risks for the companies themselves. In this paper, we want to fill this gap with two contributions. First, we present the obstacles to RAI implementation in the Swiss business landscape that we identified through qualitative interviews. Second, we develop the *Control-Tangibility Framework* that can support companies in defining priorities and timeframes for their RAI implementation plans. This framework considers both a company's ability to influence change (which we term *organizational control*) and its capacity to observe aspects of each challenge (which we term *tangibility*).

To our knowledge, this is the first work inquiring into Swiss companies' implementation of RAI. Our study provides a systematic and data-driven approach to understanding and addressing RAI implementation challenges, enabling strategic planning and thus contributing to narrowing the gap between ethical frameworks of AI and their application. Compared to the existing ethical frameworks of AI, our framework differs fundamentally in both purpose and foundation. Current ethical frameworks are normative and prescriptive—establishing ethical principles and their operationalization. Differently, our Control-Tangibility Framework serves a complementary meta-level function: rather than establishing and prescribing ethical principles, it provides a systematic methodology to prioritize and address barriers to RAI adoption. Therefore, it complements existing ethical frameworks of AI by helping organizations navigate the barriers to adopting any ethical framework they choose. Furthermore, the Control-Tangibility Framework is empirically grounded in real implementation challenges faced by organizations.

The paper is organized as follows: in Section 2, we briefly review the studies that examined organizations' implementation of RAI. In Section 3, we describe the research framework and methods of our study. In Section 4, we present the findings of our research and develop our framework of the RAI implementation challenges. In Section 5, we provide guidelines for the development of an implementation plan addressing such challenges, and we indicate the limits of our study and directions of future research. In Section 6, we conclude with our final remarks.

II. Related work

Many RAI studies focused on theoretical frameworks rather than practical challenges (for a comprehensive review, see [9]). The practical implementation of these frameworks is in its early stages and incomplete [13], [19], [20]. As highlighted by [8], existing implementation tools and methods address only selected phases of the AI lifecycle and specific RAI principles, particularly explicability. These implementation approaches often prove challenging to apply in real-world settings and demand significant technical expertise from users [8]. In addition, as they focus on the AI lifecycle, they tend to overlook how AI interfaces (and reshapes) the operations and processes of organizations. For these reasons, the translation of RAI principles into practical tools and actionable processes that can be effectively applied remains one of the main challenges of RAI.

Several studies investigated how companies implement RAI principles through various methods: surveys, interviews, and analysis of the companies' publicly available documents. Morley and colleagues found that AI practitioners faced uncertainty about the ethical alignment of AI products and difficulty implementing ethics frameworks due to their abundance and vagueness. Their study included 54 survey responses and 6 semi-structured interviews with UK-based AI practitioners [16].

In 2020, Ibáñez and Olmeda conducted 22 interviews and 2 focus groups with top and senior managers in Spanish companies to investigate how companies approached ethical issues of AI and apply RAI principles [11]. The study revealed that participants saw a disconnect between ethical principles and their practical application. Also, participants emphasized that these principles need to be tailored specifically to different sectors, types of applications, and individual projects.

Agbese and colleagues interviewed 10 Finnish executives of small enterprise software companies on their consideration and implementation of RAI requirements [15]. The authors found that middle-higher level management suggested to increase the importance of RAI requirements, and thus facilitate their implementation, by including principles like technical robustness and safety in the risk requirements, and principles like societal and environmental well-being in the sustainability requirements.

Tidjon and Khomh examined the implementation of RAI principles in 14 countries through the publicly available documents of organizations [17]. They identified five implementation gaps: lack of implementation tools, lack of effective standards, lack of training courses, weakness of the implementation of RAI principles in corporate governance, and lack of coverage of ethics in artificial general intelligence by implementation materials. Their identified mitigation strategies were fostering inclusiveness and strengthening public-private partnerships.

In conclusion, previous work reveals that current implementation tools for RAI remain constrained in their scope, primarily focusing on isolated lifecycle phases and specific principles. In most studies across multiple countries, a pronounced gap between theoretical principles and practical application is the recurring challenge in RAI implementation. To our knowledge, no study has so far developed a system that maps the RAI implementation challenges and offers an approach to systematically limit them. In this contribution, we provide insights into a previously unexplored geographical and business context, the Swiss business landscape, and introduce an empirically grounded framework for evaluating and addressing implementation challenges through careful prioritization and measurement.

III. Methods

We investigated the RAI implementation approaches of Swiss organizations in a pilot project, using qualitative methods to surface and map issues due to the project's exploratory nature and limited duration. More in details, we followed a two-step research approach. In the first step, we carried out 11 online interviews in English with professionals working directly in RAI-related positions (n=9) across Swiss organizations and academic experts who advise Swiss companies on RAI implementation (n=2).

The interviewees came from two boutique ethics consultancies, a Swiss office of a global consulting firm, a major data services company, a nonprofit, a university, an AI compliance provider, and a state-owned enterprise. Their roles

ranged from technical positions to senior management and CEO level. The aim of the interviews was to understand interviewees' firsthand experiences and viewpoints on the implementation of RAI.

In the second step of the project, we hosted an on-site workshop called "Consulting in RAI: How to narrow the RAI implementation gap" held at the University of Zurich in September 2024. This workshop, facilitated by the first three authors, brought together 13 participants—nine from six Swiss companies and four from academia. The workshop served to validate and enrich our interview results and start developing practical solutions for RAI implementation challenges.

To reach companies across diverse sectors for the interviews and workshop, we created a targeted list of Swiss organizations using our professional networks, web searches, and expert opinions and then contacted them directly. We additionally published a formal call for participation on the Swiss Innovation Agency's website dedicated to the Innovation Boosters, which was also posted on LinkedIn by the first and third authors.

Both the interviews and workshop were video recorded and transcribed. Two interviews were coded independently by the first and third author. After comparing the coding, the first author coded the remaining interviews and workshop transcripts using MAXQDA 2024 software. We analyzed both the interview and workshop data through thematic analysis.

IV. Results

A. Organizations' pain points in implementing RAI

On the basis of the thematic analysis of the interviews and workshop, we identified five areas in which organizations face issues in implementing RAI, which we call pain points: economic, structural and procedural, conceptual and technical, cultural and behavioral, and regulatory and compliance. We chose a naming convention that groups the RAI implementation challenges into high-level categories that can accommodate further challenges we may identify in the future through quantitative methods or interviews of a broader sample of companies. We present the five pain points in detail in the following.

1) Economic pains

The most frequent challenge that emerged is economic. Companies often do not have a dedicated budget for RAI and/or underestimate the costs and time required ("*It's not a project that you do in a couple of weeks, you bring in some experts, you get it done, and now you're a RAI company*," Participant (P) 5). Furthermore, as stressed by an interviewee, implementing RAI may generate indirect costs in the form of foregone profits due to the limitations that RAI requirements impose on some projects ("*certain checks and regulations could potentially impact the viability of certain projects. If I were a business manager with a strong desire to see a project through, I might consider ways to circumvent such checks*," P8).

2) Structural and procedural pains

Structural and procedural pains encompass both formal organizational structures and processes affecting RAI implementation. In this area, the three most prominent challenges are: identifying the person responsible for the RAI initiatives and their budget allocation, securing top management endorsement of RAI practices, and establishing processes to address potential trade-offs with business goals. Usually responsibility for RAI is dispersed across legal departments, IT, and digital responsibility teams, with little engagement from top management. This is problematic because top management support is generally recognized as an important success factor for organizational change [21]. In addition, RAI, like many socially-oriented goals in organizations [22], can easily be in conflict with short-term business goals. As one interviewee remarked, when top management does not consider RAI as a priority, RAI is put aside or overlooked in favor of business goals: "*if there's not a top down decision and support from the management to really do this [implement RAI], then it will not be very sustainable*" (P9). Another interviewee mentioned the need to devise processes that surface and offer ways to address such trade-offs ("*This [trade-off] raises the question of how we can streamline the process for dealing with such situations*," P8).

3) Conceptual and technical pains

Conceptual and technical pains address the dual difficulty of conceptual understanding and technical implementation of RAI into metrics and recommendations that are actionable within an organizational context. Interviewees said that they are not always sure about the meaning of privacy, transparency, trustworthiness, and responsibility and how to implement these concepts in the day-to-day operations of their organization ("*There's a real proliferation of principles and manifestos and cartas and documents [on RAI]. [...] But what does it mean in your daily work? [...] It's really difficult to tell people what [RAI] means and to operationalize it*," P9). Especially technical staff was concerned about the difficulty to find metrics for measuring the trustworthiness of AI projects. Management staff from consultancies mentioned the struggle in applying RAI principles to specific solutions for their clients. Moreover, an interviewee contended that some companies underestimate the multifaceted nature of RAI, deriving a false sense of confidence from limited RAI implementation ("*there are some people who think we're just going to slap on some fancy privacy enhancing technology and it's all fine*," P6).

4) Cultural and behavioral pains

Cultural and behavioral pains include organizational values, attitudes, and behavioral norms affecting RAI adoption. In our interviews, we found that companies have a hesitant attitude toward RAI implementation due to both the human psychological tendency and a trend of Swiss business culture of "wait and see" ("*it's humans itself with their hesitation to wait and see what is happening. But this is normal human behavior, especially in Switzerland to just wait and see what the others are doing*," P2). This wait-and-see attitude might also be linked to the considerable uncertainty in the regulatory landscape, as we discuss in the next point.

5) Regulatory and compliance pains

Regulatory and compliance pains reflect the external regulatory pressures and compliance requirements that organizations must navigate. The interviewees saw RAI implementation as strongly connected to regulations, and the uncertainty in the regulatory environment as a challenge to RAI implementation at the moment of the interview. Some interviewees contended that the EU AI Act is vague and does not provide enough guidance to companies in specific sectors ("*we know that [the EU AI Act] is still vague, particularly for certain industries. When you look at the framework of the EU*

AI Act, it is horizontal and geared toward customer protection and individual human rights. So, it's not what our financial services companies can really benefit from," P7). Several interviewees also expressed uncertainty about the upcoming regulations in Switzerland ("*it is not yet clear whether we will have to ensure compliance with the EU AI Act or whether a Swiss version of the Act will be available*," P8). As our interviews were conducted between June and September 2024, we believe the prominence of the EU AI Act in the interviewees' responses reflects interviewees' concerns about its imminent implementation and potential impact on the Swiss market.

B. *Mapping RAI implementation challenges: the Control-Tangibility Framework*

To help organizations mapping and addressing their RAI implementation challenges, we elaborate the Control-Tangibility Framework. The latter is a thematic mapping of the organizations' pain points grounded on two fundamental dimensions: organizational control and tangibility. The rationale of choosing control and tangibility as primary dimensions in classifying RAI implementation barriers is data-driven. These dimensions emerged through our interpretative analysis of the interview and workshop data and allowed us to systematically map the various challenges organizations face in RAI implementation in a way that, as we will discuss in Section V, offers valuable insights for practitioners. We provide below the definitions of the two dimensions.

- **Organizational control**: The degree to which an organization can influence, modify, or overcome a pain point in a short time frame and without depending on external actors or factors. This dimension helps identify over which pain points companies have more direct control versus those that require adaptation to external conditions.
- **Tangibility**: The property of a pain point to be directly observable through quantitative (e.g., opportunity costs) or qualitative (e.g., job descriptions) assessments. Tangibility is a crucial driver of organizational action, with both positive and negative effects. On the positive side, high levels of tangibility allow organizations to observe the effects of the initiatives they put in place, and to adjust action accordingly. In addition, individual decision makers can demonstrate success and be rewarded accordingly. Thus, organizations tend to tackle tangible challenges preferentially. However, on the negative side, nontangible challenges are often as important, if not more so, than tangible ones, but might be neglected because of the difficulty of demonstrating any effect. Including tangibility in our dimensions allows organization to surface and discuss both tangible and intangible issues.

We observed that some interviewees intuitively recognized varying degrees of organizational control and tangibility among different challenges. The Control-Tangibility Framework systematically categorize and visualizes these dimensions specifically for RAI implementation, thereby transforming implicit organizational knowledge into explicit insights that can then be discussed and for which action plans can be developed. Furthermore, our framework enables organizations to identify RAI pain points that allow for easy wins (high control and high tangibility), while also supporting the identification and visibility of pain points where actions might be less easy or direct, and discuss ways to address them.

On the basis of the interviews and our thematic analysis, we discussed the location of the pain points in a Cartesian plane in which the axes correspond to the dimensions of our framework (Fig. 1). We provide here initial guidance on the likely location of each pain point. The precise placement of the pain points in the framework may vary depending on the organization's specific characteristics and sector, which the organization should explore through collaborative discussions guided by our framework.

Economic pains. They are the most tangible challenges as the costs of RAI implementation can be estimated with some degree of certainty, as well as the time invested and the opportunity costs (in terms of risks and predictions). Economic pains are the second most controllable challenges as an organization can control budget allocation, resource prioritization, and investment decisions, but it cannot affect the conditions of the market, the competitive pressure from other companies, and regulation which might result in heavy fines.

Structural and procedural pains. These involve tangible and intangible aspects, and as a result have intermediate tangibility levels. The tangible aspects are related to the definition of formal authority, responsibility, and budget within organizations (e.g., organizational charts, role descriptions, budget allocation process). The intangible aspects are mostly linked to the informal aspects of organizing, including the extent of top management informal authority. Structural and procedural pains are the most controllable RAI pain points, since organizations can directly affect role assignments and responsibilities, as well as department structures and internal processes like budget allocation mechanisms. What organizations have less possibility to affect is the informal influence and decision-making power.

Conceptual and technical pains. These have the second lowest level of tangibility and a moderate level of control. Their degree of tangibility is due to the fact that some components of RAI are difficult to measure, especially when long-term consequences are considered. An example is autonomy of individuals in an organization. Autonomy is a key ethical principle enabling people to freely make decisions; it can be threatened by subtle strategies manipulating employees' decision making, which are often difficult to observe and measure.[2]

Other ethical components of RAI such as fairness, explainability, and transparency of AI systems can be assessed via technical metrics [23]. It is noteworthy that the choice and implementation of these metrics are normative decisions that necessitate precise definitions of the principles they are designed to represent, which may be constrained by industry standards and best practices. Addressing these pains requires the collaboration of interdisciplinary teams, such as legal,

[2] We acknowledge that the extent to which individuals are encouraged to make autonomous decisions might vary depending on the cultural factors within a company environment, but in the conceptual and technical pains we consider autonomy *qua* ethical principle.

compliance, data analytics, as well as the allocation of resources, which may involve acquiring third-party services. A further challenge belonging to this pain point is how to handle the trade-offs among different elements of RAI, for instance privacy versus transparency, as well as between RAI and business objectives. This often depends on social and legal factors that are only partially under the organizations' control. As a result, we posit that conceptual and technical pains are moderately controllable by organizations.

Cultural and behavioral pains. These are typically characterized by the lowest level of tangibility and a moderate level of organizational control. While there are tools such as questionnaires that can help companies assess and influence cultural dimensions, culture remains a relatively difficult concept to operationalize (tangibility dimension) and influence (control dimension). Changing cultures is possible, but it typically requires consistent action over long-time scales, resulting in only limited organizational control. Furthermore, an organization's size affects the control degree of cultural and behavioral pains. This is because larger organizations may be able to exert some influence on some cultural and behavioral norms of a society through communication campaigns, sponsorship, and other marketing strategies, even though the prevailing cultural and behavioral norms of a society are largely beyond their control. Differently, smaller companies do not have such an influence.

Regulatory and compliance pains. These are moderately tangible and largely uncontrollable. The moderate tangibility stems from combining a mix of tangible elements (i.e., specific law requirements with associated outcomes) with nontangible ones (i.e., interpretations of guidelines, applications to specific contexts, conjectures about the upcoming Swiss regulation). Regulatory and compliance pains are largely uncontrollable: although organizations can develop internal compliance policies, these must align with external regulations, over which organizations have limited control. The control degree may be influenced by organization size. Large organizations can potentially shape regulations through lobbying, while smaller ones typically must operate within existing regulatory frameworks.

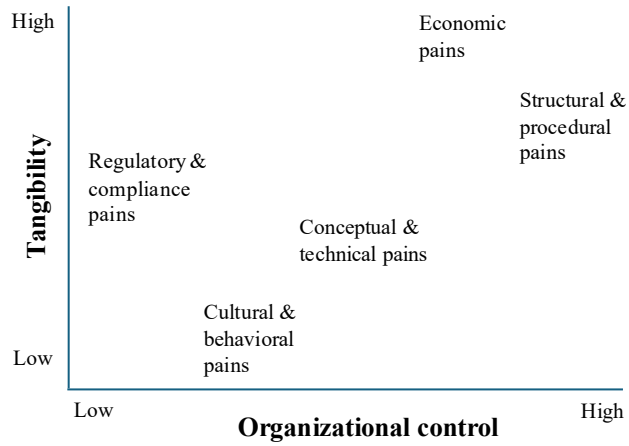

Fig. 1. Diagram showing the Control-Tangibility Framework mapping the five pain points on two axes: tangibility (how much a pain point is directly observable) and organizational control (how much influence organizations have over the pain point). This positioning visually represents how much observable and controllable each pain point is.

V. Discussion—Toward a RAI implementation toolkit

It is unlikely that organizations can implement RAI effectively by using one-size-fits-all solutions, even when these are industry specific. Organizations need to tailor their RAI approach to their specific circumstances. The Control-Tangibility Framework provides organizations with a starting point for addressing the major pain points in RAI implementation in a targeted way. We conceive our framework as the first element of a toolkit for the implementation of RAI in organizations.

In this section, we first show the value of the Control-Tangibility Framework for helping organizations crafting targeted implementation strategies. Second, we lay the groundwork for the development of a RAI implementation toolkit, of which our framework is the foundation.

A. The first RAI toolkit element: The Control-Tangibility Framework and its advantages

Our framework offer organizations several advantages to start addressing the major pain points of RAI implementation.

First, the dual-axis classification provides a conceptual map that enables organizations to assess pains through two-fold lenses. The control spectrum of our framework helps organizations to assess their own ability to influence different pains. The tangibility spectrum helps companies to elaborate appropriate indicators for different tangibility levels, using categories such as, for instance, low, moderate, and high, whereas numeric indicators might be misleadingly precise. Furthermore, the differentiation of tangibility degrees within the components of the pain points helps explain why purely technical or structural solutions often fall short: they typically address only the tangible aspects of a pain point while overlooking its less tangible but equally important aspects.

Second, jointly considering tangibility and control aspects allows a more nuanced understanding of implementation challenges, which can inform both strategic planning and operational execution. For instance, while economic pains are highly tangible and controllable, cultural challenges, though less tangible, require strategic consideration given the long-time scales and consistency of actions needed to alter cultures.

Third, the dimension spectrums facilitate the creation of comprehensive implementation plans that address both tangible and intangible aspects and prioritize interventions on the basis of the organization's ability to effect change. The progression from most to least controllable/tangible helps organizations develop staged implementation approaches, potentially increasing the likelihood of successful RAI implementation plans. In particular, the control spectrum helps companies to balance effort between controllable and uncontrollable factors and set appropriate expectations for different types of change initiatives in terms of results and timeframes. This is particularly valuable given the finding that companies often do not have a dedicated RAI budget and/or underestimate the costs and time involved.

It may be objected that companies employing our framework may be induced to focus on highly controllable and tangible pain points as they are easier to achieve than those with low control and tangibility. We reply that our framework contributes to avoid this situation by enhancing the visibility of challenges with low control and low tangibility and proposing mechanisms to address them. Furthermore, other factors contribute to motivate organisations to address

these challenges once they have been made visible. Specifically, since pain points are interconnected, concentrating on easy wins alone will result only in partial implements. In turn, partial implementation is unlikely to be sufficient for regulatory compliance and long-term business sustainability.

Overall, the value of the Control-Tangibility Framework lies in providing a structured methodology to prioritize RAI implementation efforts, creating a common language for discussing diverse RAI implementation challenges, and enabling strategic planning by highlighting which challenges require different approaches.

B. Developing further elements of the RAI toolkit

Our study also suggests directions for the development of a RAI implementation toolkit. We showed that the RAI pain points cover a wide number of organizational domains, regulatory and compliance issues that affect risk management processes, economic issues that reflect the impact on markets and revenues, and cultural and behavioral elements. The diverse nature of the pain points underscores that implementing RAI requires a comprehensive organizational transformation. A RAI toolkit will thus benefit from drawing on the extensive knowledge developed over the past decades on how to manage transformational organizational change both in general [24], [25], and specifically in relation to the digital transformation [26]. This work can provide tools and processes that can be used in the context of RAI, such as how to build momentum around change by constructing a shared RAI vision, creating supporting coalitions, and setting up networks of change champions and processes associated to the needs of digitally informed strategies, changing cultures, and similar.

However, some elements of RAI are specific and need dedicated insights and tools. These are grouped primarily in the conceptual and technical and economic pain points. In the case of conceptual and technical pains, advances in overcoming them partly depends on advances in the overall understanding of how AI ethical challenges manifest and can be addressed. Such a progress can in turn inform an organization's specific approach and development of specific tools. A first priority should therefore be the development of tools that allow the integration of the advances in research on measuring, for instance, transparency, robustness, and risk, into the existing processes and tools of organizations. This "translational" research should aim to help companies in identifying and making the connections explicit between AI adoption, the pain points that we identified, and the places where RAI issues manifest in organizations. It requires fostering collaborations with research institutions to bridge the conceptual and technical gaps that hinder the effective implementation of RAI initiatives. Organizations may initially adopt generalist tools, such as Z-Inspection®, to achieve a level of contextualization appropriate for their RAI initiatives, rather than relying on bottom-up customizations [27]. Frameworks like IBM AI Fairness 360 (https://aif360.res.ibm.com/), Google's What-If Tool (https://pair-code.github.io/what-if-tool/) and methods such as SHAP [28] serve as practical starting points, enabling organizations to operationalize RAI principles effectively. Part of this effort will require developing tools that assess the economic impact of RAI, in terms not only of risk but also of opportunities. Thus, for addressing the economic pains, a RAI implementation toolkit should include clear return-on-investment measurements for RAI initiatives. In addition, specific actions that may help companies address these aspects are, for instance, the allocation of a dedicated RAI budget and the integration of RAI costs into project planning cycles.

For the aspects of cultural and behavioral pains that are less controllable, such as the wait-and-see attitude, companies can act through change management levers that address both informational and motivational aspects. From an informational point of view, a more proactive internal culture can be fostered through targeted employee training, cross-functional collaboration, and open discussions about challenges and needs across roles. From a motivational point of view, behavior that is coherent with the new culture needs to be rewarded consistently over time across various aspects of organizational life [30]. For instance, engaging and promoting RAI initiatives needs to be considered and rewarded in promotion decisions. By harmonizing these efforts, companies can build resilience and adaptability to the cultural and behavioral pains, which are only partially controllable.

With regards to structural and procedural pain points, a particularly critical aspect concerns what employees can do to secure top management endorsement of RAI practices—an issue that RAI shares with other challenges involving ethics in business. In such a case, employees might find ways to frame RAI issues in terms more congenial to organizational goals and values, such as risk mitigation and competitive advantage. Additionally, employees could build cross-functional alliances with colleagues in legal and compliance departments who might share their concerns.

To address challenges beyond an organization's control, companies are encouraged to develop a contingency plan. This plan should outline the actions to be taken in case of unanticipated events that are out of an organization's control. Such events may be triggered by, for example, unexpected or rapid shifts in regulatory requirements, or the introduction of new cultural values into business. A contingency plan requires elaborating a business impact analysis, an incident response plan, a recovery plan, and a business continuity plan [29]. In practice, there are different methodologies for developing and implementing each component. While thorough development of each component is essential, it is equally important that the various components, once formulated, are continually refined to ensure a state of readiness for any potential incident.

C. Limitations of the study

The Control-Tangibility Framework has limitations due to our pilot project constraints. Our study's sample size was limited in both the number of interviews conducted and companies examined, potentially affecting how well it represents all needs and pain points across different company roles. Additionally, since the study focused exclusively on Swiss organizations, the findings primarily reflect the business environment in Switzerland. While these insights offer valuable perspectives on the Swiss context, they cannot be automatically generalized to other countries without conducting similar studies that account for the specific business conditions, regulatory frameworks, and cultural factors in those regions. As a consequence, our Control-Tangibility Framework may exclude some pains that were not in the interview data. Furthermore, the clear-cut categorization along control and tangibility spectrums might not accurately reflect the nuanced reality where challenges often exist on a

continuous spectrum rather than in discrete categories. For instance, the distinction between "most controllable" and "moderately controllable" might be more fluid in practice than the Control-Tangibility Framework suggests. Similarly, the framework's generalizability across different organizational contexts, industries, and scales may be limited. For instance, industry-specific factors might alter the relative tangibility of certain challenges and what constitutes a highly controllable challenge in a large organization might present differently in a smaller one, as we have previously contended.

D. Future research directions

The limitations of our pilot project suggest several avenues for future research. First, longitudinal studies could help understand how implementation challenges evolve over time and how their position on the control and tangibility spectrums might shift during implementation. Second, comparative studies across different organizational contexts could help refine the generalizability of the Control-Tangibility Framework. In addition, these studies could provide empirical evidence to clarify the relationship between the tangibility and control levels of organizational challenges, on the one hand, and the costs organizations incur to address them systematically, on the other. This, together with additional research into the dynamic interactions between challenges, could lead to more sophisticated models that better capture the complexity of RAI implementation challenges. In this regard, in the future we plan to partner with Swiss organizations to develop targeted models of RAI implementation challenges and RAI implementation tools addressing these challenges.

VI. Conclusion

RAI frameworks and implementation tools aim to ethically align AI systems. Currently, organizations struggle to use both [11], [14], [15], [16], [17]. In this study, we investigated why companies cannot operationalize RAI and provided indications for developing a RAI implementation plan.

Our study makes two key contributions to understanding and addressing RAI implementation challenges. First, through a qualitative methodology, we identified five critical pain points that organizations face when implementing RAI: economic, structural and procedural, conceptual and technical, cultural and behavioral, and regulatory and compliance. Our findings reveal that while RAI principles and frameworks continue to proliferate, organizations face significant practical challenges in translating them into operational practices. The study's focus on Swiss organizations provides valuable insights into RAI implementation within this specific business context.

The second contribution of our study is the Control-Tangibility Framework, which offers organizations a structured approach to understand and prioritize RAI implementation challenges. By mapping challenges along the dimensions of control and tangibility, organizations can develop more effective implementation strategies that account for both their ability to influence change and capacity to observe and measure aspects of the challenges. Our framework thus contributes to bridging the gap between RAI principles and their implementation. Also, as AI technology continues to advance and regulatory frameworks evolve, our framework can adapt to organizations' specific contexts and needs.

Looking ahead, our research lays the groundwork for more targeted studies on RAI implementation strategies and opens new avenues for developing implementation plans to bridge the gap between RAI principles and their operational implementation. While our pilot study has limitations in terms of sample size and geographical scope, it provides a foundation for future research on RAI implementation across different organizational contexts and jurisdictions.

A successful RAI implementation requires organizations to move beyond viewing it as merely a compliance exercise and instead integrate it as a fundamental aspect of their AI design, development, and deployment processes. Our empirical findings and framework provide a practical starting point for organizations to begin this crucial transformation.

Acknowledgments

We are deeply grateful to our interviewees and workshop participants for dedicating their valuable time to this research. Editorial support for this paper was provided by Claude from Anthropic.

References

[1] M. Loi, C. Heitz, A. Ferrario, A. Schmid, and M. Christen, "Towards an ethical code for data-based business," presented at the 2019 6th Swiss Conference on Data Science (SDS), IEEE, 2019, pp. 6–12.

[2] Google, "AI Principles – Google AI." Accessed: Feb. 05, 2025. [Online]. Available: https://ai.google/responsibility/principles/

[3] High-Level Expert Group on AI, "Ethics guidelines for trustworthy AI," 2019. Accessed: Feb. 03, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[4] L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," Machine learning and the city: Applications in architecture and urban design, pp. 535–545, 2022.

[5] B. C. Stahl, "Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems," Scientific Reports, vol. 13, no. 1, p. 7586, 2023.

[6] V. Dignum, "Ensuring Responsible AI in Practice," in Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, V. Dignum, Ed., Cham: Springer International Publishing, 2019, pp. 93–105. doi: 10.1007/978-3-030-30371-6_6.

[7] M. Sadek, E. Kallina, T. Bohné, C. Mougenot, R. A. Calvo, and S. Cave, "Challenges of responsible AI in practice: scoping review and recommended actions," AI & Society, pp. 1–17, 2024.

[8] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices," Science and Engineering Ethics, vol. 26, no. 4, pp. 2141–2168, 2020.

[9] V. Vakkuri, K.-K. Kemell, and P. Abrahamsson, "AI ethics in industry: a research framework," arXiv preprint arXiv:1910.12695, 2019.

[10] J. Krijger, T. Thuis, M. de Ruiter, E. Ligthart, and I. Broekman, "The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations," AI and Ethics, vol. 3, no. 2, pp. 355–367, 2023.

[11] J. C. Ibáñez and M. V. Olmeda, "Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study," AI & Society, vol. 37, no. 4, pp. 1663–1687, 2022.

[12] R. Eitel-Porter, "Beyond the promise: implementing ethical AI," AI and Ethics, vol. 1, no. 1, pp. 73–80, 2021.

[13] R. Ortega-Bolaños, J. Bernal-Salcedo, M. Germán Ortiz, J. Galeano Sarmiento, G. A. Ruz, and R. Tabares-Soto, "Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems," Artificial Intelligence Review, vol. 57, no. 5, p. 110, 2024.

[14] J. Zhou and F. Chen, "AI ethics: From principles to practice," AI & Society, vol. 38, no. 6, pp. 2693–2703, 2023.

[15] M. Agbese, R. Mohanani, A. Khan, and P. Abrahamsson, "Implementing ai ethics: Making sense of the ethical requirements," presented at the Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, 2023, pp. 62–71.

[16] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, "Operationalising AI ethics: barriers, enablers and next steps," AI & Society, pp. 1–13, 2023.

[17] L. N. Tidjon and F. Khomh, "The different faces of ai ethics across the world: a principle-implementation gap analysis," arXiv preprint arXiv:2206.03225, 2022.

[18] B. Friedman, "Value-sensitive design," interactions, vol. 3, no. 6, pp. 16–23, 1996.

[19] J. Ayling and A. Chapman, "Putting AI ethics to work: are the tools fit for purpose?," AI and Ethics, vol. 2, no. 3, pp. 405–429, Aug. 2022, doi: 10.1007/s43681-021-00084-x.

[20] P. B. de Laat, "Companies committed to responsible AI: From principles towards implementation and regulation?," Philosophy & Technology, vol. 34, no. 4, pp. 1135–1193, 2021.

[21] S. H. Appelbaum, S. Habashy, J. Malo, and H. Shafiq, "Back to the future: revisiting Kotter's 1996 change model," Journal of Management Development, vol. 31, no. 8, pp. 764–782, 2012.

[22] T. Ramus and A. Vaccaro, "Stakeholders matter: How social enterprises address mission drift," Journal of Business Ethics, vol. 143, pp. 307–322, 2017.

[23] OECD.AI Policy Observatory, "Catalogue of Tools & Metrics for Trustworthy AI." Accessed: Feb. 05, 2025. [Online]. Available: https://oecd.ai/en/catalogue

[24] J. P. Kotter, "Leadership change," Harvard Business School Press: Boston, MA, USA, 1996.

[25] D. Buchanan et al., "No going back: A review of the literature on sustaining organizational change," International Journal of Management Reviews, vol. 7, no. 3, pp. 189–205, 2005.

[26] G. Westerman, D. Bonnet, and A. McAfee, Leading digital: Turning technology into business transformation. Harvard Business Press, 2014.

[27] D. Vetter et al., "Lessons learned from assessing trustworthy AI in practice," Digital Society, vol. 2, no. 3, p. 35, 2023.

[28] S. Lundberg, "A unified approach to interpreting model predictions," arXiv preprint arXiv:1705.07874, 2017.

[29] P. Clark, "Contingency planning and strategies," presented at the 2010 Information Security Curriculum Development Conference, 2010, pp. 131–140.

[30] J. A. Chatman and S. E. Cha, "Leading by leveraging culture," California Management Review, vol. 45, no. 4, pp. 20–34, 2003.