Managing Ethical Risks of Artificial Intelligence in Business Applications

Quintin McGrath[1], Alan R. Hevner[1], and Gert-Jan de Vreede[1]

[1]Affiliation not available

February 27, 2024

Abstract

The introduction of artificial intelligence (AI) capabilities in business applications provides significant benefits but requires organizations to manage critical risks of AI ethical consequences. We survey a range of large organizations on their use of enterprise risk management (ERM) processes and toolsets to predict and control the ethical risks of AI. Four serious gaps in current ERM systems are identified from analyses of the survey results: (1) AI ethical principles do not translate effectively to ethical practices; (2) Real-time monitoring of AI ethical risks is needed; (3) ERM systems emphasize economic not ethical risks; and (4) When ethical risks are identified, no solutions are readily at hand. To address these gaps, we propose a proactive approach to manage ethical risks by extending current ERM frameworks. An enhanced ERM (e-ERM) framework is designed and evaluated by subject matter expert focus groups. We conclude with observations and future research directions on the need for more aggressive pro-ethical management oversight as organizations move to ubiquitous use of AI-driven business applications.

I. Introduction

The effective integration of Artificial Intelligence (AI) technologies and methods into business applications generates significant benefits for organizations and their stakeholders through efficiency gains, greater repeatability, and new business models [1]. However, these benefits must be balanced with the potential for unintended ethical consequences resulting in business and stakeholder harm as seen by the many examples found in AI incident databases.11E.g., The AI Algorithmic and Automation Incidents and Controversies repository which is found at: https://www.aiaaic.org/home, The AI Incident Database which is found at: https://incidentdatabase.ai/, or AI Global's dataset which is found at: https://map.ai-global.org/ AI technologies have the potential to revolutionize business applications not only because of their unique capabilities (e.g., self-learning potential, intelligent capabilities, and inscrutability) but also because of AI's integration and interdependence with human users, which establishes it into tight socio-technical systems of human-plus-machine [2-4].

The management controls of AI-based business systems design, implementation, and use provide significant challenges. While some initial control approaches have been offered [5, 6], we focus on risk management in an organization, particularly its processes and tools for enterprise risk management (ERM). Wirtz et al. [7], in defining an integrative framework for the governance of AI, identify six dimensions of AI risks: "(1) technological, data, and analytical (2) informational and communicational, (3) economic, (4) social, (5) ethical, as well as (6) legal and regulatory AI risks" [7]. To constrain the scope, rather than attempt to address all these risk dimensions, our research focuses on the dimension of ethical risks associated with AI-based solutions (AIS). An organization's ethical climate and culture support a set of ethical principles and norms that its leaders espouse and lead to the ethical practices that staff members are held accountable

to [8]. These AI ethical principles may be applied to AIS creation and use [9, 10]. Deviating from these ethical principles and norms creates a risk for the organization.

Our research goal is to study the current state of business organizations to manage ethical risks in AIS. We present a survey of large organizations on their use of enterprise risk management (ERM) processes and toolsets to predict and control the ethical risks of AI. From analyses of the data, we identify four critical gaps in current ERM systems: (1) AI ethical principles do not translate readily to ethical practices; (2) Real-time monitoring of AI ethical risks is needed; (3) Most ERM systems focus on economic not ethical risks; and (4) When ethical risks are identified, no solutions are at hand. To address these gaps, we propose a proactive approach to manage ethical risks by extending existing ERM frameworks. An enhanced ERM (e-ERM) framework is designed and evaluated by subject matter experts. We conclude with observations on the need for more aggressive pro-ethical management oversight over AI-driven business applications.

II. Background

There is a high level of international visibility and, thus, attention to the regulation of AI-based applications. For example, the USA National Institute of Standards and Technology's (NIST) AI Risk Management Framework version 1.0 [11] is a guidance document for managing and mitigating the risks associated with creating and operating AIS. The goal of the framework is "to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems" [11]. Further, the European Union's AI Act aims to ensure safe and trustworthy AI applications using a "regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI" [12].

The International Standards Organization 23894:2023 standard extends the ISO 31000:2018 Risk Management standard [13] to address AI-related risks by adding AI-specific guidance. It has similar goals to the NIST AI RMF, to guide the management of risk related to the creation and use of AI. It goes further to "assist organizations to integrate risk management into their AI-related activities and functions" and describes the processes necessary for the implementation of AI risk management [14].

A. Enterprise Risk Management

Organizational risk can generally be described as the uncertainty of an event and its outcomes, which could harm the performance of an organization or the achievement of its goals [15, 16]. Risk is a function of an event, its likelihood of occurring, and the anticipated extent of the impact should it occur. Risk management is a process that considers risks and opportunities associated with an action. It is successful when the greatest potential benefit is achieved with the least possible harm. According to Hillson [17], risks are first identified through various structured and unstructured processes within the *risk management cycle* . Second, risks are assessed to determine the likelihood of their occurrence and potential impact. The assessment results in a matrix of probability versus potential impact. The third step in the risk management cycle defines the actions that should be taken to maximize benefit over harm. These include avoiding, transferring, limiting the impact or likelihood, or accepting the risk and planning for the potential impact. Action can also be deferred due to time constraints or lack of a good solution. [13, 15, 17]

ERM applies this concept to the whole enterprise and may be understood as a business strategy process and set of tools used to achieve business goals [18]. This will lead to organizational and stakeholder value maximization for an organization. The risk categories associated with ERM are organization-dependent but can generally be categorized into financial, operational, strategic, and compliance risks [15, 18]. Others also include reputational risk [e.g., 19]. We propose that ethical risks should be an additional ERM risk category, especially when AIS are created and used in an organization.

B. Artificial Intelligence Systems

As a working definition for this research, we define AI as *the ability of an algorithm to perform tasks commonly associated with humans.* We further sharpen the focus on the cognitive abilities of AI that Sheth et al. [20] describe as the "ability to simulate human thought process in a computerized model." These cognitive abilities result in "cognitive services," which include AI capabilities related to language, speech, vision, knowledge, decision support, and search [20, 21]. The Organization for Economic Co-operation and Development (OECD) illustrates this integration and interdependence in their *Framework for the Classification of AI Systems* [22]. They define five dimensions that can be grouped into the AIS dimensions of the "AI model," its "Data and Input," and its "Task and Output." These are related to the two context dimensions in which it operates, the "Economic context" and the "People and Planet" dimensions. These dimensions interact and influence one another within the socio-technical solution.

A cognitive AIS socio-technical solution is particularly complex as the technical aspect of the AI (e.g., self-learning potential, intelligent capabilities, and inscrutability) may continue to change, and the environment in which it operates may also vary in unpredictable ways. NIST posits that as socio-technical systems, AI systems "are influenced by societal dynamics and human behavior" [11]. They go on to state that the risks and benefits of the AIS can, therefore, emerge from how an organization uses it and the social context in which it is used. Its socio-technical nature makes AIS a complex technology both for businesses and their stakeholders.

Seeber et al. [4] indicate that human and AI interaction in a socio-technical system is more than a human accessing an intelligent algorithm; it requires communication, acting together as a team, and for the AI to be aware of the human (and we would add the human's awareness of the AI), especially when this happens in an uncertain and unpredictable environment. With the current pace of AI advances, they indicate that the ethical and moral challenges are being elevated, implying the need for us to grapple with the questions of whose ethics need to be considered by the algorithm and the level of AI's agency. Further, Asatiani et al. [2] indicate that the operating logic that AIS employs in approaching problems may differ from how humans approach the problem, making it hard to understand and explain the AIS's decisions or actions. As it learns, the AIS may exhibit biases based on the data it is trained on. This becomes more complex when the AI system uses self- or reinforcement learning.

C. Artificial Intelligence Ethics

The field of ethics can be described as consisting of: (1) meta-ethics, which considers universal moral truths and the nature of right and wrong; (2) normative ethical theories that relate to the principles and standards used to judge right from wrong; and (3) applied ethics, which are the practices used when faced with ethical dilemmas [23, 24]. This study does not consider meta-ethics in detail, it simply accepts the existence of a universal moral truth and adopts the concepts of right and wrong.

Ethical principles, specifically AI ethical principles, are core to this research. They may be considered "part of a normative theory that justifies or defends moral rules and/or moral judgments; they are not dependent on one's subjective viewpoints" [25, p. 365]. Furthermore, these AI ethical principles are derived from the universal ethical principles that an organization inculcates as its organizational code of ethics. Applied ethics are the principles utilized in solving ethical dilemmas or making ethical decisions. AI ethical practices is an example of applied ethics.

Ethical risks are the unexpected negative consequences of those actions misaligned with an organization's norms or codes of ethics. Francis [26] stated that "ethical risk is to be seen as a part of overall risk management" and that "managing ethical risk is an important aspect of managing risk in general." Thus, we argue that ethical risk is an additional ERM category that must be managed as a core part of the enhanced ERM framework. When these AI capabilities are considered in the light of ethical theory, it is possible to derive a set of AI Ethical Principles to guide a business's design, development, and use. We adopt Floridi

and Cowls' [27] AI Ethical Principles of autonomy, beneficence, non-maleficence, justice, and explicability for this study.

III. Industry Survey on Ethical Risks

To investigate the current use of ERM to manage AI ethical risks in large business organizations, we performed a large-scale online survey using Qualtrics11https://www.qualtrics.com. Questions addressed the extent of using AI-enabled solutions across organizations, approaches to AI ethics, the related AI risks, and how they were managed. We used an online structured survey approach, using Prolific22https://www.prolific.co/, to gather the information. For us to obtain a large enough sample, online panels were used. It was necessary to use two surveys due to the nature of the panel consisting of respondents from larger organizations creating AI-enabled solutions based in the United States. The first screened potential participants. We then invited the filtered respondents to participate in the second survey.

The use of an online panel provided two advantages. First, it yielded a random sample outside our influence. Second, it allowed a blind study in which we were aware of neither the organization nor the name of respondents (only their unique identifier). This permitted greater privacy for the respondents and their organizations than we could otherwise have provided.

A. Survey Questions

Saunders et al. [28] enumerate several well-known research strategies for research, including surveys. Tengli [29] indicated that the survey method was helpful in analyzing data collected from a sample of participants to answer what, who, and where types of research questions. When considering the research strategies available, we were guided by the overall research question. We used surveys to gain a broad understanding of what was happening in the industry. The design for the surveys was informed by Hewson [30] and Toepoel [31], with the questions in the survey built on findings from Corea [32], Morley et al. [10], and Shneiderman [6], among others.

The survey consisted of several sections: First, the **consent**form was followed by questions on the **demographic information**used to validate the responses and filter out poor inputs. It included questions relating to the role of the responder, to confirm that they had an AI decision-making role, and to the use of AI-based resources. The third section tested the **importance of AI** to the organization and asked about the level of investment in AI-enabled solutions. The fourth and fifth sections as about the organization's**experience in AI** and the **extent of the implementation** . The latter was tested by asking which organizational components it was deployed to, the AI capabilities used [32], and the anticipated benefits [33]. The next section asked about the organization's**approach to AI ethical risks** through a request to rate a selection of risks from the literature [34-38]. Their use of risk approaches [39] was then assessed.

The survey asked about **AI ethics** in terms of, first, the organization's approach to its **AI Ethical Principles** [36, 38] and, second, whether these principles had been inculcated into AI Ethical Practices that were implemented and enforced. AI Ethical Principles are those used in the e-ERM framework [40]. In addition, the frequency of use of the underlying ethical lenses was assessed, namely, fairness/justice [41, 42], duty [43], virtue ethics [42, 44], common good [42], utilitarian [42, 44, 45], and rights [42, 44, 45]. The level of the implementation of the AI Ethical Principles was probed, considering if they were "Documented," "Published," "Training provided," or "Enforced and monitored."

A selection of **AI Ethical Practices** [6, 33] was then provided, and the responder was asked to indicate the extent of use and efficacy, and, if they were not present, were they desired ("Available, used, and beneficial," "Available and used," or "Not available, but desired" [33]). As with the AI Ethical Principles, the extent of the implementation was assessed ("Documented," "Published," "Training provided," "Embedded in daily practices and processes," and "Enforced and monitored").

B. Survey Execution

We issued the Qualtrics survey via Prolific's website to a predefined panel of respondents who indicated that they were currently employed in organizations in the United States and who had been active on Prolific in the last 90 days. This represented a pool of 128,904 potential responders. After an initial pilot test of 20 participants to ensure that it was working as planned, we opened the survey to groups of 20, 30, 50, 100, and 200 based on capacity to process and analyze the responses.

We received a total of 2,047 screening survey responses over a 6-week period. Of the respondents whose organizations were developing AI-enabled solutions, 69 organizations employed less than 100 people. These were removed as they did not meet our selection criteria of a 100+ person organization. The next filter eliminated respondents with no responsibility for AI, even if their organizations were developing AI solutions. Those without this responsibility were considered too remote from the AIS development process to provide accurate responses to the main survey. As a result, 45.3 percent of the balance (230 respondents) were eliminated, resulting in a candidate pool of 278 responses. One response was excluded in error, resulting in 277 respondents who we invited to participate in the main survey.

We piloted the main (second) survey with a group of doctoral students to ensure that the questions were clear, that the logic worked correctly, and that the results were being properly processed by Qualtrics. We made minor changes to the wording of some questions and to the spacing and page breaks to make the survey flow more easily.

Part way through the survey, we changed a single question relating to the most critical AI-related risks. After the first 100 responses, it was clear that the question was not producing clear results. It was therefore adjusted to force a selection of the three most critical AI risks instead of asking the participants to rate all the risks on a 5-point Likert scale. This latter approach resulted in an unambiguous response from the remaining respondents.

The 277 custom panel respondents were invited to participate in this main survey. This was done using a "custom allow list" containing only the selected Prolific IDs. The survey was issued via the Prolific web site in the same manner as the screening survey. The main survey ran for two months and resulted in 206 valid responses.

C. Demographics of Survey Participants

Comparing the demographics of the respondents to the 2019 data from the U.S. Census Bureau [46], it was evident that the sample was not representative of the distribution of the population of all organizations with more than 100 employees in the United States (see Table I). For instance, there were significantly fewer smaller organizations in the sample than would be required to be representative and significantly more responses from larger organizations. Even so, the responses did provide a balanced sample across the various organization sizes, notwithstanding some overrepresentation of the 1,000- to 9,999-size organizations and under-representation in the 10,000- to 19,999-size organizations.

From a statistical validity perspective, the number of responses for each organizational size group was sufficient to provide a margin of error of 7.6 percent or less for a 95 percent confidence level across the groups (see Table I). So, conclusions can be drawn from the screening survey results about the level of AI-enabled development in each organizational size segment. It was possible to compare their levels of AIS development. For the entire sample, the margin of error was 2.2 percent at the 95 percent confidence level, which implied that we could draw general conclusions from the data, including, for instance, in which parts of the organization AI-enabled solutions were generally deployed and the representative AI-related roles of the various leaders and staff within an organization.

TABLE I: Organizational Size – Margin of Error

	less than 100	100-499	500-999	1,000-9,999	10,000-19,999	More than 20,000
Population of organizations (using 2019 US Census)	5,986,587	94,957	10,348	9,398	582	540
Sample	608	408	250	381	130	266
z-value for 95% confidence	1.96	1.96	1.96	1.96	1.96	1.96
Margin of Error	4.0%	4.8%	6.1%	4.9%	7.6%	4.3%

IV. Survey Results

In the screening survey, 28.2 percent of the respondents indicated that their organizations are developing AI-enabled solutions. When normalized for all organizations, 11.6 percent of all organizations are developing AI-enabled solutions and about 26.8 percent of those organizations with more than 100 staff are doing so. As we anticipated, the results reveal that the percentage of organizations developing AI-enabled solutions increased with the size of the organizations indicating that larger organizations can more effectively invest in the development of AIS. This expectation is confirmed with a linear regression of the percentage of those organizations developing AIS for each of the organizational size ranges sampled, which resulted in a positive slope coefficient of 0.072, an R^2 of 0.938 and a p-value of 0.001, thereby demonstrating a strong linear correlation with a high certainty of validity.

A. AIS Investments and Anticipated Benefits

Considering the importance and investment levels in AIS in the various organizations, we find that the likelihood of organizations to create AI-enabled solutions increases from around one in four for organizations in the 100-499 range of employees to more than 50 percent for organizations with over 20,000 staff. This implies that the larger the organization, the greater propensity for AI-enabled solutions.

The main survey provided more insights into the importance of AI to the organizations. All 206 responses to this survey indicated that organizations developing AI-enabled solutions consider AI to be more than moderately important. The mean of the responses was 2.675 (which is above the mid-point of 2.5 with a range of 1 to 4) with a kurtosis of -0.756, implying that more respondents select *very-* or *extremely-important* than those who select the *moderate-* or *slightly-important* options. In terms of spending on AI-enabled solutions as a proportion of spending on other assets, based on 192 responses, the mode is 10-20 percent of the overall asset budget, with a smaller percentage of responses above 20 percent than those below 10 percent of the overall asset budget. The mean is 2.667 (below the mid-point of 3 with a range of 1 to 5) with a kurtosis of -0.701, thus leaning slightly towards less than 10-20 percent of asset expenditure on AIS development.

The position that AI was more than moderately important for most of the participating organizations was supported by an average spend in the 10 to 20 percent range for the organizations. At face value, the importance of AI drove the level of spend on average across the responding organizations. So, those organizations that believed that AI-enabled solutions were important for their organizations, spent more on AI (see Table II). For those organizations with greater than 20,000 staff, 72.2 percent rate AIS as very important or extremely important compared to 60.2 percent for all 206 responses, with higher spend mean of 2.830.

TABLE II: Importance of AI vs. Percentage of Capital Spend on AI

Importance vs. Spend	Count	Less than 5%	5 to 10%	10 to 20%	20 to 50%	More than 50%
Slightly Important	25	59.10%	4.50%	27.30%	9.10%	0.00%
Moderately Important	57	20.80%	34.00%	35.80%	7.50%	1.90%
Very Important	84	4.94%	25.93%	44.44%	23.46%	1.23%
Extremely Important	40	0.00%	30.60%	33.30%	33.30%	2.80%

Note : the average spend percentage for that level of importance is shown as a progressive bar

The main anticipated benefits from AI-enabled solutions are improved competitiveness, improved profits, increased market opportunities, and greater product quality (see Table III). There are also high levels of agreement on the benefits of AIS for improved reputation, enhancing customer trust, improving staff satisfaction, and improved customer loyalty. The lowest agreement is for greater social impact, but it is still seen to be a benefit by more organizations than those who do not.

TABLE III: Anticipated Benefits

Anticipated Benefits	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
Improved competitiveness	87	85	26	5	3
Improved profits	82	88	28	4	2
Increased market opportunities	73	97	28	5	2
Greater product quality	76	90	31	8	1
Improved reputation	48	88	56	7	7
Enhanced customer trust	43	96	49	14	3
Improved staff satisfaction	42	96	49	13	6
Improved customer loyalty	43	87	59	13	4
Improved staff retention	34	68	69	23	12
Greater social impact	34	66	62	27	15
Other	9	6	44	2	7

Note : n=206

B. AI Risks

The survey samples the most concerning risks associated with the use of AI. The responses are contextual as they differ organization to organization, differ based on the AIS application used, and differ on the cognitive capabilities applied. Overall, the answers provide interesting insights regarding the general level of concern, and, thus, the priority with which the risks should be mitigated. There are 105 responses to this question with each survey participant allowed to select up to three risks they are most concerned about (see Fig. 1).

The AI risk of highest concern is lack of accuracy or reliability of the AIS (selected by 44 of the 105 participants), followed, 7.6 percentage points lower, by concerns over the lack of control of the AIS by humans. Safety, security and robustness, fear and trust, lack of accountability, and lack of transparency are grouped together as the next most concerning set of AIS-associated risks.

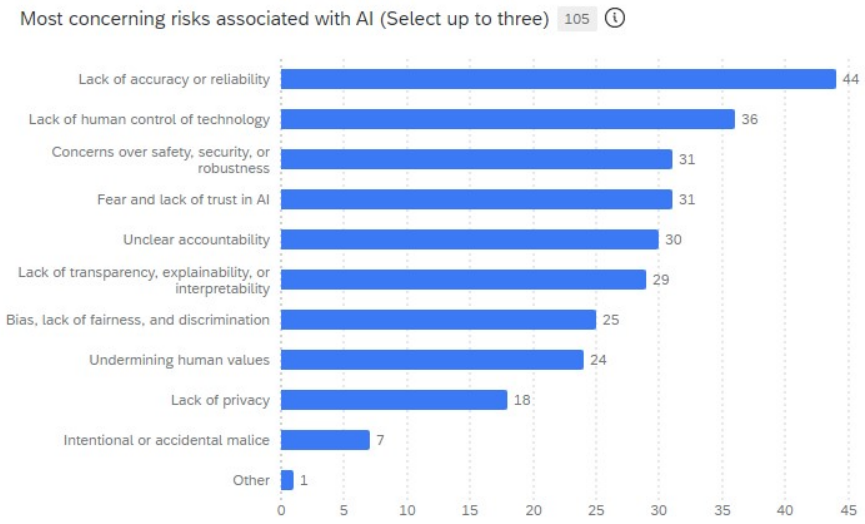


Fig. 1 . Most Concerning AI Risks (n-105, up to three selections permitted)

C. Maturity of AI Ethical Principles

While many organizations document (69.9 percent) and train their staff on the AI Ethical Principles (56.8 percent), there are fewer who focus on monitoring and enforcing the use of these principles (41.7 precent) or making them available by publishing them (36.9 percent). (see Fig. 2).

Fig. 2. Implementation of AI Ethical Principles (n-201, multiple selections permitted)

D. Maturity of AI Ethical Principles

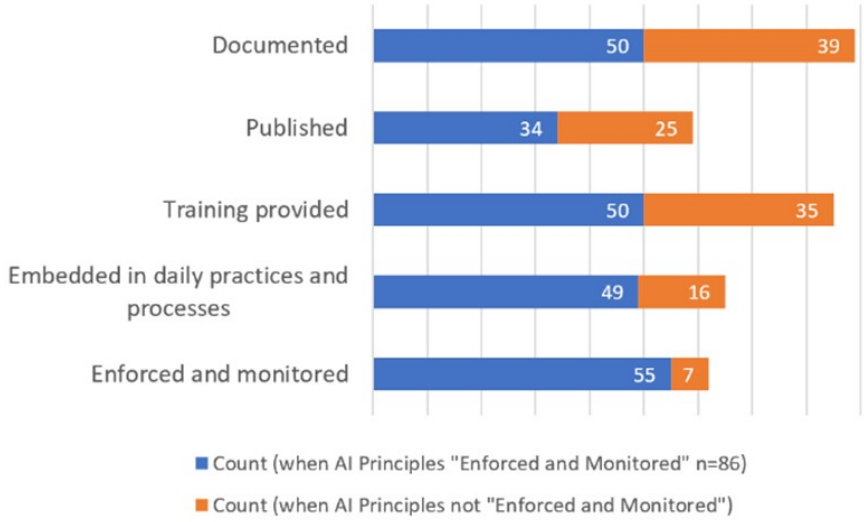


Fig. 3 provides insights into the level of definition of AI Ethical Practices in the responding organizations. A total of 57.8 percent of the respondents indicate that their AI Ethical Practices are defined. A total of 89 out of the 119 organizations with defined AI Ethical Practices also have the practices documented, 59 have them published, and 85 provide regular training. Of the 86 organizations that have AI Principles enforced and monitored, 55 of them (64 percent) also have AI Practices enforced and monitored.

E. Risk Approaches Used

Another important research goal is to gain an understanding of the risk approach used by the responding organizations. The survey showed that all but 13 of the 206 respondents have a structured way of addressing risk management. The majority (150 organizations) make use of their technology processes, 75 organizations use their existing overall ERM processes, and four use another process (e.g., the use of a cross-functional AI review board, spot check validations, and tests run in selected groups). Respondents can select multiple answers, indicating that 33 organizations use both ERM and technology processes, two use technology and another process, and one use ERM and another process.

We anticipated that the larger the organization, the more prevalent the use of ERM for AI risk management would be, but this did not turn out to be the case (see Table IV). For the three organizational size groups below 10,000 people this is the case, with a linear regression of R^2 of 0.977 (p -value of 0.098), but when all groups are included, this moves to R^2 of 0.309 (p -value of 0.33). Thus, there is no clear relationship between the larger size of the organization and its greater propensity to use ERM for AIS risk management. There is also no significant relationship between organization size and the use of technology risk processes. It was apparent, as anticipated, that the smallest organizations (100-499 employees) have a greater chance of not using a structured risk management approach.

TABLE IV: Distribution of the Risk Management Approaches Compared to Organization Size

Org Size	ERM	Technology Processes	Another	None
100-499	28.3%	67.4%	0.0%	13.0%
500-999	34.0%	78.7%	4.3%	4.3%
1,000-9,999	44.1%	78.0%	0.0%	3.4%
10,000-19,999	38.9%	66.7%	5.6%	5.6%
more than 20,000	36.1%	69.4%	2.8%	5.6%

V. Gap Analysis of Ethical AI Enterprise Risk Management

A gap analysis identifies and measures the space between 'where we are' (present state) and 'where we want to be' (desired state) in the assessment of a business unit. Gap analyses are used in many disciplines to support the identification of critical problem gaps and the magnitude of the challenges required to close the gaps to achieve satisfactory solutions. For example, Robinson et al. [47] apply gap analysis to identify research challenges in clinical healthcare and Lawn et al. [48] present a gap analysis for gambling rehabilitation research.

The results of the literature-based survey above were used to perform a gap analysis of the present state of ERM management of ethical AI risks and a desired state of proactive AI ethical awareness and control. We identified four significant gaps, each of which is elaborated below along with the related analysis.

A. Gap 1 – AI Ethical Principles -> AI Ethical Practices

Our survey results demonstrate that most organizations formally document AI Ethical Principles with required training requirements. However, the translation of principles to practices is inconsistent. The lack of clear AI ethical practice statements results in confusion and limited enforcement when ethical principles are violated. Many of the organizations do not monitor or enforce ethical practices (just 62), and only 65 have embedded the AI Ethical Practices into the organization's daily operational processes. Nearly two-thirds of the organizations that monitor and enforce the principles, have well-defined ethical practices in place, while only 57.0 percent have ethical practices embedded in the organization's daily processes and practices.

This gap between the definition and implementation of AI Ethical Principles in an organization and the moving of these principles from their cognitive acceptance to practical implementation in the day-to-day business activities of the organization is a major concern. If all the organization has is a set of principles that do not change its daily practices, they are of little practical effect. This "principles-to-practice" quandary is a recognized problem in the literature, and while many researchers have proposed solutions that include frameworks, toolkits, and guidelines [5, 6, 49-51], their lack of organizational adoption persists [33, 52-54].

Unfortunately, AI Ethical Principles are often abstract, complex, and challenging to apply in practice by those implementing AIS and making them tangible and implementable may be difficult [33, 54, 55]. For instance, translating the AI Ethical Principle of "autonomy" (e.g., the level to which decision-making is allocated from humans to the algorithm) into practice when creating an AIS requires us to answer questions like what determines which decisions or types of decisions are retained and which are delegated to AI? How and when should we revoke the delegation to the artificial agent? Designers, developers, and testers of AIS require specific rules and guidelines on how to address the need for autonomy in the system. It is often unclear who in the organization has the responsibility to translate AI Ethical Principles into practices to guide their work. Without these practices, they find it hard to be sure that the autonomy principle is effectively implemented within their AIS. While addressing this quandary may be left to the AIS designers,

developers, and testers to solve, we and others [e.g., 34, 56, 57] argue that it is, in fact, a broader challenge that business and IT leaders should be accountable for.

B. Gap 2 – Static Assessment of AI Systems -> Dynamic Monitoring of AI Systems

The survey data provide a rich understanding of how organizations' view AI risks in Figure 1. An analysis of the highest ranked risks (e.g., accuracy, human control, security, accountability, transparency, etc.) point to a need for real-time monitoring of the AIS performance and decisions. The nature of these risks requires a dynamic approach to risk management which ERM static risk assessments of AI systems do not address. This gap is exacerbated by the costs and complexities of performing real-time monitoring of critical business applications with time-sensitive requirements.

Inherent to an AIS is the potential of making decisions, providing recommendations, or arriving at conclusions using a logic that is not clear to the user. Second, considering our working definition of AI, namely, "the ability for an algorithm to perform tasks commonly associated with humans," acting like humans implies the potential for integration and interdependence between AI and its human users. Because this is at an "intelligent" level, it establishes a tight socio-technical system of human plus machine [2, 3]. Further, this complex socio-technical system operates within a multi-faceted environment of dynamic social expectations, developing regulatory rules, and ever-changing cultural perspectives. These complexities create the need for increased business risk management agility and responsiveness.

As a result, fixed-cycle ERM approaches with scheduled reviews fall short. To maximize the business and societal gain and to minimize harm, an agile and dynamic ERM approach is needed. This requires a continuous process because of the dynamic nature of the environment in which an organization operates. ISO [13] defines eight principles of risk management to enable value creation and value protection. One of these principles is the dynamic principle, described as follows: "Risks can emerge, change or disappear as an organization's external and internal context changes. Risk management anticipates, detects, acknowledges and responds to those changes and events in an appropriate and timely manner" [13, Section 4[e]].

This need for agility is core to a business's ability to sustain competitive advantage, and leaders of an organization achieve agility by sensing and responding to opportunities and risks, seizing opportunities, and making appropriate changes to the organization [58]. Nair et al. [59] focuses on ERM as a dynamic capability that enables businesses to react and respond positively to change in the environment. Bogodistov and Wohlgemuth [60] suggest that ERM needs a dynamic capability to allow for the continuous reassessment of risks to drive strategic and operational risk responses [60]. This dynamic capability is especially important in relation to high velocity changes caused by technologies such as AI. An organization's business leadership will need to address this gap through making the ERM approaches more dynamic and responsive.

C. Gap 3 – ERM focuses on Economic Risk -> ERM balances Ethical and Economic Risks

The survey demonstrates that organizations value broader benefits from AIS, like reputation, customer trust and loyalty, and social impact. This requires that the ERM focus must be more than purely economic. Risk categories associated with ERM depend on the organization but fall into the broad categories of financial, operational, strategic, and compliance risks [15, 18]. Others [e.g., 19] have added reputational risk to this list. Wirtz et al. [7] define a broad classification of AI risks using six categories, one of which is AI ethical risks.

This need to include an ethical aspect into the ERM approach is critical because of the socio-technical nature of AIS. Above, we discussed AIS acting like humans, leading to the potential for integration and interdependence between AI and its human users. This leads to the need to define and adhere to AI Ethical Principles like those defined by Floridi and Cowls [27] in terms of autonomy, beneficence, non-maleficence,

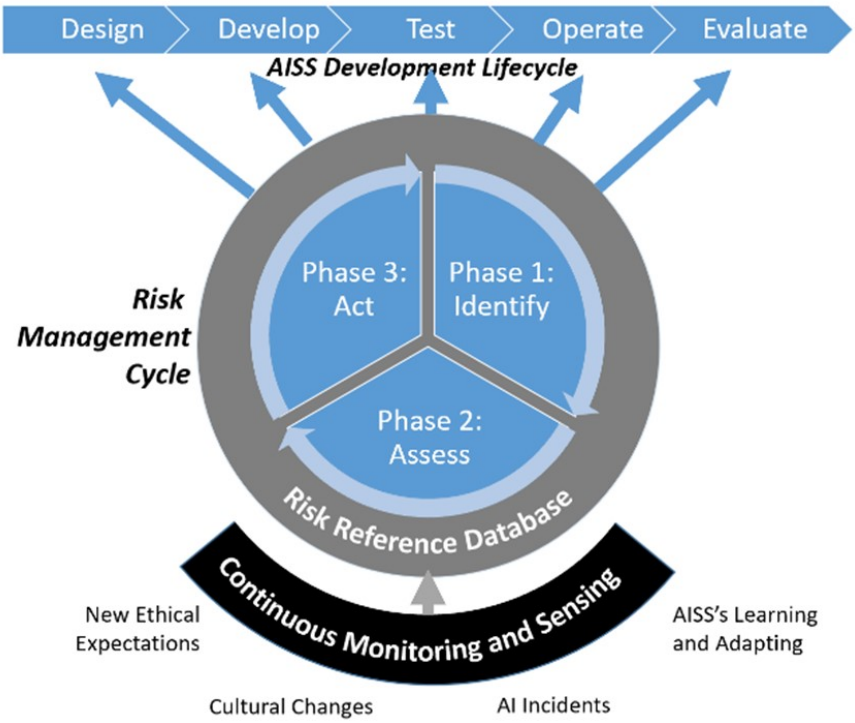
justice, and explicability. Violating these principles and norms leads to ethical risks which need to be balanced with the appropriate financial risks. Thus, we must extend ERM processes to place greater weight on ethical AI risks to achieve an effective and sustainable balance with economic risks. In this way, the ERM will continue to drive the achievement of organizational goals and value maximization but do so in an ethical way.

D. Gap 4 – Ethical Risk Problem Identified -> Ethical Risk Solutions Designed and Implemented

While concept of ERM is well understood, there is a gap in its application in an integrated way with technology risk processes to provide solutions to mitigate the ethical risks associated with the AIS applications. The survey shows that all but 13 of the 206 respondents have a structured way of addressing risk management. The majority (150 organizations) use their technology processes, 75 organizations use their existing ERM processes, and four use another process (e.g., a cross-functional AI review board, spot check validations, and tests run in selected groups). Respondents were able to select multiple answers, indicating that only 33 organizations used both ERM and technology processes, two used technology and another process, and one used ERM and another process. The data indicate a wide range of approaches for addressing ethic risks once identified. However, in all cases there is a 'hand-off' to another internal group to design and implement a solution to ethical risk found. Thus, there is a 'solution' gap made more evident when dynamic monitoring (Gap 2) reveals the presence of a significant ethical risk and no 'at-hand' remedy is available. This reinforces the need for an ERM framework in which available solutions are designed beforehand and ready for application.

VI. Extending ERM for AI Ethical Risks

Business leaders who are most effective, assess opportunities and their associated risks using ERM frameworks [17, 61]. These frameworks use the risk management cycle to maximize business and stakeholder value by first identifying risks, second, assessing the probability of the occurrence of a positive or negative risk event, and third, acting to mitigate it. AI-enabled solutions are complex, adding both new risks and new opportunities because of their capabilities and the socio-technical nature of the AIS.



AIS operate within uncertain and unpredictable environments. Environmental dimensions include changing ethical expectations, like emerging public pressures relating to facial recognition [62, 63], a developing regulatory environment exemplified by the recent release of the European Union AI Act [12], and an evolving cultural environment as the system is used in new countries around the world [64]. As a result, standard risk management approaches fall short. To maximize business and societal gain, a more agile and dynamic ERM approach is needed. Fig. 4 presents a conceptual model of an enhanced ERM (e-ERM) process that is embedded in both the AIS development life cycle and as a monitoring capability in the performance of AIS applications. The enhanced model is compatible with existing ERM frameworks from the ISO 31000:2018, Risk Management Process [65], and the NIST conceptual AI Risk Management Framework [66]. For instance, the NIST model recognizes the need to integrate risk management into the AIS lifecycle (i.e., pre-design, design and development, test and evaluation, and deployment). This integration between risk management and the AIS development lifecycle drove the need for managing risks throughout the creation of the AIS and once it is operational.

VII. e-ERM Extensions

We now provide further extensions to e-ERM to address the four critical gaps. Fig. 5 provides an expanded vision of the e-ERM with the proposed extensions.

A. AI Ethical Principles to Practices

Organizations must ground AI ethical principles with actionable practices that relate to business applications. We recommend the development and use of a *risk reference database* (RRD) to record the evolving risks and the best practices to mitigate them. The RRD can be grounded on evolving experiences both internal and external to the organization. Based on the literature, many frameworks, toolkits, and guidelines have linked AI Ethical Principles to proposed AI Ethical Practices. These included the IEEE's Ethically Aligned

Design principles [5] for AI system design, Shneiderman's [6] fifteen proposals for bridging the gap between ethics and practice using an enhanced governance approach, Peters et al.'s [50] proposal for balancing a responsible design process with the user's interaction with the technology, and Vakkuri and Kemell's [51] nine-step "RESOLVEDD strategy" for rational ethical decision-making. These various approaches form the foundation for part of the problem space that the RRD can address.

The RRD will relate the AI Ethical Principles to relevant AIS capabilities, which would guide the AIS creators to select the most pertinent AI Ethical Practices for their solution. In addition, specific AI capabilities appear to be more controversial than others, e.g., facial recognition. The RRD could take specific AI capabilities and applications and relate them to the most applicable AI Ethical Principles, which could highlight the most appropriate AI Ethical Practices. This filtering process would reduce the need to consider all the possible AI Ethical Practices and limit them to a relevant subset based on the AI capabilities and applications being considered.

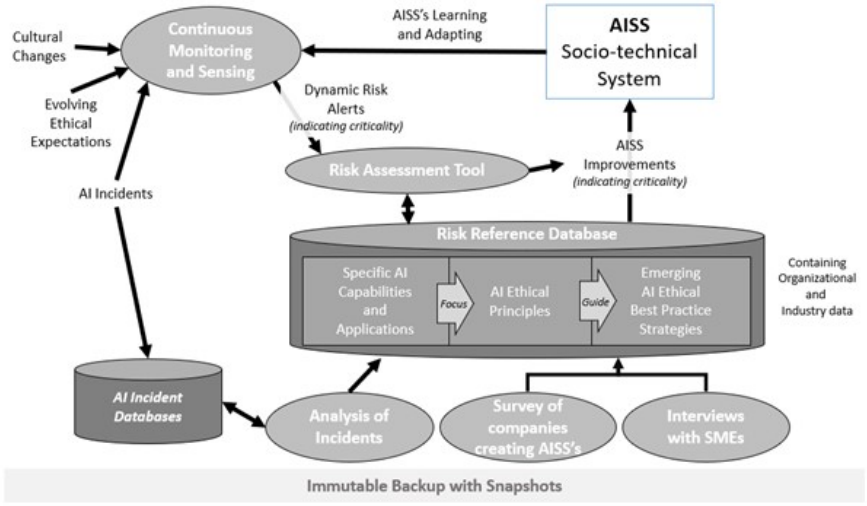


Fig. 5. Extensions to e-ERM to Manage AI Ethical Risks

Segmentation and extension of the RRD for industry and organization types are essential. Also, because of the need for a point-in-time state to be maintained, e.g., in the case of litigation, an immutable, secure (role-based) backup solution for the e-ERM is needed.

B. Dynamic Monitoring of AIS Applications

To support real-time reactions to occurrences of AI risks, we identified three essential components: mechanisms for continuously monitoring and sensing (CMS component), an agile risk assessment tool and approach (RAT), and the aforementioned risk reference database (RRD) to record the evolving relationships between risks and the best practices to mitigate them.

Because of the changing nature of the AIS and its operating environment, effective monitoring of the operation of the AIS by the CMS is necessary. The problems the CMS needs to address includes identifying emerging risks in both direct and indirect environments. Manual and AI agent-based tracking are required to register changes in system behaviors and deviations from defined goals (e.g., biased decisions, privacy exposures, and low transparency of outputs). From an environmental perspective, active operational monitoring is needed to identify any changes in outcomes (e.g., failures due to unforeseen inputs, exploitation of vulnerabilities, and activities of adversaries). In terms of the broader environment, the analysis of new AI incidents is needed to establish new insights into the evolving ethical expectations of the AIS users. The evolution of cultural perspectives on ethics is also required to be monitored.

Based on its monitoring and sensing, the CMS needs to trigger risk alerts with an indication of the criticality to initiate the risk management cycle and an underlying risk assessment tool (RAT) capability. The RAT component would support the execution of the risk management process described above. Importantly, it must not be isolated from the rest of the organization's ERM risk assessment process but integrated to ensure that AIS-related risks are appropriately handled along with the other enterprise risks.

Another vital part of the RAT concerns engaging the right people at the right time in various risk assessment steps. Nagbøl et al. [67] segments the risk assessment process into three distinct modules that each relate to a specific group. This allows them to target the right participants achieving an efficient risk assessment. The first module considers business needs for the AIS, which engages business domain experts. The second focuses on the system's technical details, thus using data scientists. The final module is a synthesis of the outputs of the previous two modules requiring a combination of business and data science experts. We use a similar approach in the e-ERM design.

C. Balancing Ethical Risks with Other Categories of Risk

ERM approaches support the accepted risk management cycle described above (see Fig. 4). These solutions address the generally accepted risk categorized into financial, operational, strategic, compliance [15, 18] and reputational risk [19]. Core to the Risk Assessment Tool (RAT) is the risk management cycle of identification, assessment, and mitigation, considering the various risk categories. For the e-ERM to be effective, an organization's core ERM should be well established and embedded into its business processes.

One way of understanding the additional impact of the AIS on the risk profile of an organization is to consider the European AI Act [68]. The Act, which is built on a legal framework of ethical principles, "combines a risk-based approach based on the pyramid of criticality, with a modern, layered enforcement mechanism" [69]. The European Union's [70] pyramid of criticality classified AI systems into levels of criticality based on the types of remedies suitable for four different levels of risk. AI systems at the unacceptable level of risk can cause substantial harm, such as cognitive behavioral manipulation, and are banned from use. High-risk systems need careful assessment and monitoring (e.g., critical infrastructure or law enforcement relating to fundamental rights). Limited-risk systems have specific transparency requirements (e.g., chatbots). Finally, minimal or no-risk AI systems are those that pose no direct risk to users (e.g., spam filters).

This range of ethical risks based on the European Union's classification (and other classifications as they develop) is an important input into the RAT of the e-ERM system because it will guide the necessary risk-mitigation actions based on the type of a proposed AI-enabled solution. Consequences of ineffective risk management within the European Union are dire in terms of fines and sanctions.

Thus, the balance of the effective management of the generally accepted ERM risks along with the incorporation of the additional ethical risks resulting from the creation and use of AIS in an organization is essential. Not doing so may result in censure and significant financial consequences to a business, and thus the proposed e-ERM design.

D. Availability of AI Ethical Risk Solutions

Our research identified that there are multiple approaches used in business today to mitigate ethical risks associated with AIS. These include making use of current ERM solutions in the organization, the use of IT processes, or a combination of these. A small number of organizations use other approaches. As indicated, the study reveals the presence of significant ethical risks for which no 'at-hand' remedies are readily available.

To address this gap, the e-ERM design proposes the use of the Risk Reference Database to link the risk identification and assessment process (using the RAT) with potential solutions through the identification of best practice mitigation approaches associated with the ethical risks. The selection of the most appropriate approach is determined by considering what AI capability is used in the system, and therefore what AI ethical risk category is impacted. For example, if the trigger is an ethical risk associated with facial recognition,

the risk category may relate to privacy, an element of the non-maleficence AI ethical principle. A range of best practices could include removal of the facial recognition capability from the AIS, constraining the use of facial recognition, or adding additional controls.

Identifying, collecting, and categorizing these AI best practice solutions is complex as organizations that have experienced AIS-related incidents are reluctant to reveal what changes they made to mitigate the issue because of the potential of litigation. Solving the free flow of emerging best practices is a topic for future research. For this study, the AI incident databases, where some organizations have indicated what was done to address the issue, provide sufficient examples to validate the design.

E. e-ERM Scenario

As a proof-of-concept demonstration that illustrates the novel extensions of the e-ERM process, we use a realistic scenario of an ethical AI incident that was reported on the AIAAIC repository. The incident related to pressure groups' response to a web conference service's planned use of AI emotional recognition to "monitor and detect the feelings and emotions of its users" [71]. This is not the first time this emotional AI capability has raised concerns, as according to the AIID, a popular recruitment system removed expression tracking due to a complaint [72].

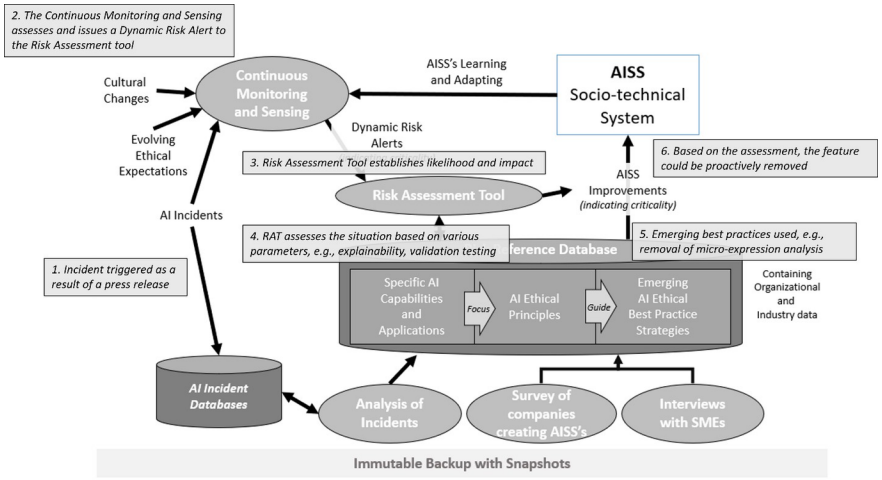


Fig. 6 . e-ERM Scenario Walk-Through

Fig. 6 uses the example of an organization that has created or deployed an AIS with emotion detection capabilities and implemented the e-ERM framework. The steps it takes are the following: (1) as the press release relating to the pressure groups' concern with the organization's plans is recorded as an incident in the AIAAIC repository, it (2) triggers the CMS, which issues a dynamic risk alert since the organization has deployed an AIS system that uses emotional analysis. Then, (3) the risk tool establishes the likelihood and impact of the risk, and (4) the RAT assesses the situation based on the information available, e.g., the reason for the concerns being raised by the pressure groups, the related level of validation testing, and the ease of explaining the organization's use of the emotion detection capability. As a result, (5) based on the emerging best practices, e.g., another organization's decision to remove micro-expression analysis from its tool, (6) the organization may choose to remove this feature from the deployed AIS or enhance the system to provide better explanations of its decisions.

This same process could be triggered by a deviation from a design goal because of the AIS's learning and adapting to a change in the cultural environment or new ethical expectations resulting from the EU [12] AI

Act. These examples require a dynamic and responsive ERM approach to ensure that benefit continues to be maximized with minimal harm for the business and its stakeholders.

VIII. e-ERM Focus Group Evaluation

We conducted two structured focus groups to evaluate the new e-ERM model with subject matter experts in ethical risk management [73, 74]. The focus group participants were drawn from experienced managers who were familiar with ERM processes. We presented the participants with the revised e-ERM model (Fig. 5) and asked them to provide feedback on the utility, effectiveness, and implementation possibilities of the model for assessing the risks of ethical AI violations.

In order to analyze the output of the focus groups, we used three coding cycles based on Saldana [75]. Because the focus groups used a structured approach of topics and broad questions, the most appropriate first cycle of coding was structural coding. The second coding cycle used axial coding to group the initial codes from the first coding cycle into conceptual categories. Finally, in line with the selected thematic structure for the focus groups, the third coding cycle extracted the themes for each of the categories. We then applied these themes to the various topics to identify those aspects that the focus group participants supported, and those aspects that needed to be changed or adapted.

We identified 47 code phrases, which were grouped into eight categories, and finally combined into three major themes. The themes were: (1) core attributes of the e-ERM model; (2) the organizational governance categories; and (3) impact factors in the broader organizational environment. The themes can be illustrated as three concentric circles with the model at the center, and the other two themes emanating from it (see Fig. 7). These were used validate the model in its ability to address the four critical ERM gaps. The categories in the outer, external environmental theme circle reinforces the importance of a clear ethical framework in which the model operates. Other environmental issues included the cultural impact on ethics, the differing ethical risks in various industries, and the speed of changes in this evolving and dynamic area of AI ethics. The focus group results reinforced those ethical risks that were pertinent to the model. They highlight how the model addresses closing gap 3 to ensure that the ethical environmental aspects are considered, as well as addressing the dynamic nature of AI risks (gap 2).

Two categories make up the governance-related issues theme. These are the pre-requisites for the e-ERM model, which related to the definition of the benefit of the system to stakeholders and employees, the goals of the model, and the foundational elements of the model. The implementation considerations related to how best to implement the model across the organization. These focus group results confirmed the need for the e-ERM model as a means of addressing gap 4.

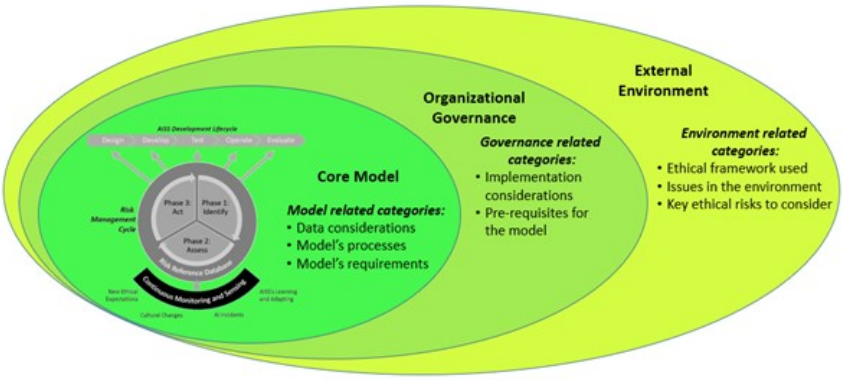


Fig. 7. Focus Group e-ERM Themes

The categories included in the model-related theme focus on three aspects, namely, clarifying the model's requirements, its processes, and data collection and maintenance considerations. These focus group results focus on the model itself (gap 4), and confirmed the processes needed to close gap 1, the AI Principles to Practices quandary.

IX. Discussion

The landscape of ethical AI risk management in business applications is continuously changing with new problem spaces and solution opportunities. Our objectives in this research are to better understand the current state of how organizations apply ERM processes to manage risk and to propose enhanced capabilities (e-ERM) to close current processing gaps. The following observations are relevant to our research findings.

A. e-ERM Implementation

The three-phase risk management cycle of identifying the risk, assessing its likelihood of occurrence and potential impact, and then acting to mitigate the risks is often executed on a regular time-based cycle, e.g., every quarter. This study, and the three recent documents, NIST AI RMF, EU AI Act, and the ISO 23894:2023 standard, highlight the need for a dynamic approach to ERM, indicating that changes in the internal and external context drive the risk changes, requiring risk management to sense and respond to the changes in an agile way. This same expectation was discovered in discussions with the focus groups. Some may contend that their organizational processes are in line with the dynamic ERM requirement from the literature, but we argue that, unless these are integrated into a broad approach like that of the e-ERM, the management of the ethical risks from AIS will be difficult to achieve.

B. AI Ethical Incident Repositories

One of the important elements of the e-ERM is the AI Incident Database. Rather than create a new database, there are a number of currently available databases. The first step in the process is the identification of publicly available databases of AI incidents. Three potential repositories are:

- AI Incident Database [76]. McGregor [77] worked with the XPRIZE Foundation and established the AIID which provides "a systematized collection of incidents where intelligent systems have caused safety, fairness, or other real world problems" [77].
- AI, Algorithmic, and Automation Incidents & Controversies Database [78]. An independent, open, public interest resource that is the most comprehensive, detailed, and up-to-date resource of its kind, the AIAAIC Repository details 850+ incidents and controversies driven by and relating to AI, algorithms, and automation. Started in June 2019 as a project to better understand the reputational and other risks of these important technologies, the repository helps researchers, academics, advocates, policymakers and industry experts across the world get a better handle on how to design, develop, deploy, and regulate them. [78]
- AI Global Database [79]. This dataset provides a record of helpful and harmful AI around the world.

For an approach to assessing the quality of the data in these AI incident databases, we turned to Cichy and Rass [80] who provide an overview of the available data quality frameworks. By analyzing the twelve common frameworks, they posit that the most frequently occurring dimensions were completeness, timeliness, accuracy, consistency, and accessibility. Wang and Strong [81] defined each of these as follows:

- **Completeness** : The extent to which data are of sufficient breadth, depth, and scope for the task at hand.
- **Timeliness** : The extent to which the age of the data is appropriate for the task at hand.
- **Accuracy** : The extent to which data are correct, reliable, and certified free of error.

- **Representational Consistency** : The extent to which data are always presented in the same format and are compatible with previous data.
- **Accessibility** : The extent to which data are available or easily and quickly retrievable. [81]

We then analyzed these databases using the above criteria. In addition, we assessed the level of metadata relevant for this research. For the completeness criteria, the focus was on the level analysis done by the database curators, the availability of links to source data, and whether the most recent incidents reaching the press were included. Timeliness was concerned mainly with how long it took for data relating to the latest incidents to be made available in the database. This was measured from the time of the incident to its appearance in the database. Accuracy relied on the level of peer reviewing provided and the validation of the information through cross references of the sources. Consistency was judged on the format of the data from record to record. Accessibility focused on the ease of access to the data and if there were any strict controls on its use for this research. The availability of metadata was judged on the presence of valuable tags and assessments. Based on analysis, we fed the information from the selected sources into a spreadsheet and a model built that formed the foundation of the design of the RRD.

C. Addressing the Principles to Practices Gap

One of the main causes of the principles to practice gap was the level of definition, documentation, monitoring, and enforcement of both the AI Ethical Principles and the related AI Ethical Practices. The main survey found that only about 53 percent of the organizations sampled had defined their AI Ethical Principles, with 42 percent monitoring and enforcing them. Around 58 percent of the respondents indicated that they had their AI Ethical Practices defined, with 30 percent monitoring and enforcing them, and a smaller number indicated that they had been embedded into the organization's daily operational processes. Without publishing, monitoring, and embedding these practices, the principle-to-practice gap will remain.

Further, embedding practices that are not perceived as beneficial was a potential second reason for the continued gap. From the main survey, several AI Ethical Practices were available for use but were not seen as beneficial. These included ethical frameworks, ethical design specifications, and explainable UI designs. While these practices are key for pro-ethical solutions, if AIS creators do not perceive them to be beneficial, they will, over time, stop using them, thereby reinforcing the principles-to-practice gap. Thus, the second reason indicated for this continued gap was practices not being perceived as beneficial and therefore being ignored over time.

A potential third cause of the gap was that many of the practices were applications of common business practices to the AIS space. For instance, building of trust and credibility is a common practice for a business leader as they seek to grow their business. What we found was that the practices did not need to be new or complex, but many were simply standard business practices appropriately applied to mitigate AI ethical risks. As a result, a portion of the principles-to-practice gap could be related to trying to define a complex AI-specific practice, when a common practice would suffice.

X. Conclusions and Future Research Directions

We conducted this research in two complementary stages. To assess the current use of ERM processes for managing AI ethical risks we surveyed over 200 large business organizations. From the analysis of the survey results, we identified four critical gaps driving the need to extend current ERM approaches: (1) AI ethical principles do not translate effectively to ethical practices; (2) Real-time monitoring of AI ethical risks is needed; (3) ERM systems emphasize economic not ethical risks; and (4) When ethical risks are identified, no solutions are readily at hand.

For the second research stage, we designed and evaluated an enhanced ERM process model (e-ERM). The novel e-ERM approach maximizes business and stakeholder value through the effective management of dy-

namic AI ethical risks presented by advanced AIS. We contribute to the ERM research by extending existing proposals for managing AI risks [6, 27, 33, 55, 57]. The results of our research contribute to practice by enriching the software engineering technical procedures at the organizational level by adding actionable risk management procedures to the safety culture and closing the AI ethical principles-to-practice gap. Through this study, we have co-created the e-ERM framework with subject matter experts and practitioners. The designed framework is ready for implementation at organizations that recognize the potential of AI-enabled solutions and are seeking to implement AIS in a way that is most beneficial for their organization and stakeholders. Of particular importance to business leaders are the resulting design principles that provide specific actions that can be taken by their AIS creators, process changes they should make, and aspects of the external environment that should be monitored and responded to. Implementing this e-ERM framework has the potential to significantly impact organizations that wish to take advantage of it.

As with any research, this study has limitations. While the e-ERM model has been evaluated in rigorous focus groups with subject matter experts, it has yet to be fully implemented in an operational organizational environment. Experimental evaluation of the e-ERM approach in context requires the consideration of operational controls and the definition of measures to assess improvements in the ethical results of AIS operations. Along these lines, we identify several important future research directions.

Research must assess the effectiveness of the dynamic e-ERM engine to identify and respond to changes in the inputs (e.g., the change in the AIS, its deployment to a new geography with a different culture, or the impact of regulatory changes). An approach would be to perform a manual audit of the system's responses to ethical triggers. Linking AI capabilities to the appropriate AI ethical practices and principles would be an important to study e-ERM effectiveness. A sample of the recommendations made by the e-ERM could be used and their effectiveness tested by subject matter experts in terms of systems or process changes.

A future research project on the e-ERM monitoring and sensing capabilities relates to AI incidents. The amount of data and the frequency of the updates on AI incidents quickly become unmanageable. We must implement and test the efficacy of using AI-enabled agents to continuously scan press releases and website posts to identify published fixes. For instance, pointing a machine learning-based AI agent at the press releases relating to incidents in the AI incident database to provide a concise summary of the incident and to identify any fixes discussed in the articles would be beneficial. Further, using AI-enabled agents to discover trends and provide early indications of emerging incident patterns would enrich the e-ERM. Emerging from this research would be a clearer understanding of the relationship between the AI incidents and the related ethics, as well as a clearer understanding of how the ethical perceptions of AI's capabilities are changing.

Beyond the AIS's internal changes, our research touched on the impact of national culture on ethics, and the monitoring of changes in culture-based ethical thinking is embedded into the current e-ERM framework. The current design can be informed by additional research into the extent of the impact of changes in ethical expectations worldwide. One of the approaches Awad et al. [82] used for this purpose is the longitudinal study called the "Moral Machine Experiment" that seeks to crowd-source moral perspectives on ethical dilemmas, like those experienced by self-driving cars, to understand the variations in ethical perspectives across countries. Additional research into the cultural implications on ethics and how they change over time, e.g., based on the work of the World Value Survey [83], will enrich the e-ERM framework.

We bounded the research to cognitive AI solutions. AI continues to evolve, and the capabilities are quickly becoming more sophisticated. With this evolution comes further ethical expectations that need to be managed carefully. A potential future research project in this space could considering how effective the e-ERM framework is at managing the risks associated with self-learning AI algorithms like Generative Pre-Trained Transformers [84]. These transformers have been trained on large amounts of data and are being used, for instance, to generate text, write articles, produce marketing collateral, and to do so without the human in the loop. The potential risks are broader and less predictable than those of cognitive AI discussed in this research, so a different risk management framework and approach may be needed. Automated mechanisms to track and respond to what the AIS is doing are therefore needed. Research on how this can be done effectively is critical.

As another topic for future research, one may also consider how ethics-based risk management research relates and generalizes to other emerging technologies like quantum computing. Quantum computing is a powerful emerging technology that makes use of a faster and more powerful approach to computing, as a result significantly accelerating AI's capabilities. In addition, its non-classical computing design, opens new opportunities to address problems particularly suited to quantum computing [85]. There are already some researchers, e.g., Kop [86], who are investigating applying AI ethical thinking to quantum computing. Many of the principles described in the e-ERM framework are applicable to ethics associated with quantum computing, so research into quantum ethics to contribute to the thinking, approaches, and regulations is valuable.

References

[1] S.-L. Wamba-Taguimdje, S. Fosso Wamba, J. R. Kala Kamdjoug, and C. E. Tchatchouang Wanko, "Influence of artificial intelligence (AI) on firm performance," *Business Process Management Journal,* vol. 26, no. 7, pp. 1893-1924, 2020, doi: 10.1108/bpmj-10-2019-0411.

[2] A. Asatiani, P. Malo, P. R. Nagbøl, E. Penttinen, T. Rinta-Kahila, and A. Salovaara, "Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems," *Journal of the Association for Information Systems,* vol. 22, no. 2, pp. 325-352, 2021, doi: 0.17705/1jais.00664.

[3] E. E. Makarius, D. Mukherjee, J. D. Fox, and A. K. Fox, "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization," *Journal of Business Research,* vol. 120, pp. 262-273, 2020/11/01/ 2020, doi: 10.1016/j.jbusres.2020.07.045.

[4] I. Seeber *et al.* , "Machines as teammates: A research agenda on AI in team collaboration," *Information & management,*vol. 57, no. 2, pp. 1-22, 2020, doi: 10.1016/j.im.2019.103174.

[5] Institute of Electrical and Electronics Engineers,*Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems* , 1 ed., 2019. [Online]. Available: https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

[6] B. Shneiderman, "Bridging the gap between ethics and practice,"*ACM Transactions on Interactive Intelligent Systems,* vol. 10, no. 4, Article 26, pp. 1-31, 2020, Art no. 26, doi: 10.1145/3419764.

[7] B. W. Wirtz, J. C. Weyerer, and I. Kehl, "Governance of artificial intelligence: A risk and guideline-based integrative framework," *Government Information Quarterly,* p. 101685, 2022, doi: 10.1016/j.giq.2022.101685.

[8] M. Kuenzi, D. M. Mayer, and R. L. Greenbaum, "Creating an ethical organizational environment: The relationship between ethical leadership, ethical organizational climate, and unethical behavior,"*Personnel Psychology,* vol. 73, no. 1, pp. 43-71, 2020, doi: 10.1111/peps.12356.

[9] A. Harrison, D. Spagnuelo, and I. Tiddi, "An ontology for ethical AI principles," *Semantic Web Journal,* 2021. [Online]. Available: http://www.semantic-web-journal.net/system/files/swj2713.pdf.

[10] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, and L. Floridi, "Ethics as a service: A pragmatic operationalisation of AI ethics," *Minds and Machines,* vol. 31, no. 2, pp. 239-256, 2021, doi: 10.1007/s11023-021-09563-w.

[11] National Institute of Standards and Technology,*Artificial Intelligence risk management framework (AI RMF 1.0)* , 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[12] (2021). *Artifical intelligence act* . [Online] Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021PC0206&from=EN

[13] International Organisation for Standardization, "ISO 31000:2018(en) Risk management — Guidelines, online browsing platform (OBP)." [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en

[14] *ISO/IEC 23894:2023(E) Information technology — Artificial intelligence — Guidance on risk management* , International Standard International Organization for Standardization, Switzerland, 2023.

[15] K. Cormican, "Integrated enterprise risk management: From process to best practice," *Modern Economy,* vol. 05, no. 04, pp. 401-413, 2014, doi: 10.4236/me.2014.54039.

[16] M. McShane, "Enterprise risk management: History and a design science proposal," *The Journal of Risk Finance,* vol. 19, no. 2, pp. 137-153, 2018, doi: 10.1108/JRF-03-2017-0048.

[17] D. Hillson, "Extending the risk process to manage opportunities," *International Journal of Project Management,* vol. 20, no. 3, pp. 235-240, 2002, doi: 10.1016/s0263-7863(01)00074-6.

[18] R. M. Steinberg, M. E. A. Everson, F. J. Martens, and L. E. Nottingham, "Enterprise risk management - Integrated framework: Executive summary." [Online]. Available: https://www.coso.org/Documents/COSO-ERM-Executive-Summary.pdf

[19] B. W. Nocco and R. M. Stulz, "Enterprise risk management: Theory and practice," *Journal of Applied Corporate Finance,* vol. 18, no. 4, pp. 8-20, 2006, doi: 10.1111/j.1745-6622.2006.00106.x.

[20] A. Sheth, H. Y. Yip, A. Iyengar, and P. Tepper, "Cognitive services and intelligent chatbots: Current perspectives and special issue introduction," *IEEE Internet Computing,* vol. 23, no. 2, pp. 6-12, 2019, doi: 10.1109/mic.2018.2889231.

[21] T. E. Marshall and S. L. Lambert, "Cloud-based intelligent accounting applications: Accounting task automation using IBM Watson cognitive computing," *Journal of Emerging Technologies in Accounting,* Article vol. 15, no. 1, pp. 199-215, Spring2018 2018, doi: 10.2308/jeta-52095.

[22] Organisation for Economic Co-operation and Development, "OECD framework for the classification of AI systems," 2022, doi: 10.1787/cb6d9eca-en.

[23] S. Bonde and P. Firenze. "A framework for making ethical decisions." Brown University. https://www.brown.edu/academics/science-and-technology-studies/framework-making-ethical-decisions (accessed 2021).

[24] T. Ewest, "The challenges within ethical leadership theories," in *Prosocial Leadership* : Palgrave Macmillan, 2018, pp. 23-42.

[25] P. Di Mattia, "Ethical principles," in *Encyclopedia of public health* , W. Kirch Ed. Dordrecht: Springer Netherlands, 2008, pp. 364-367.

[26] R. D. Francis, "Ethical risk management," in *Global Encyclopedia of Public Administration, Public Policy, and Governance* , A. Farazmand Ed.: Springer International Publishing, 2016, pp. 1-5.

[27] L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," *Harvard Data Science Review,* vol. 1, no. 1, pp. 1-15, 2019, doi: 10.1162/99608f92.8cd550d1.

[28] M. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students* , 8th ed. Harlow, England: Pearson Education Limited, 2019.

[29] M. B. Tengli, "Blog 132-Research Onion: A Systematic Approach to Designing Research Methodology," in *Agricultural Extension in South Asia* vol. 2022, ed, 2020.

[30] C. Hewson, "Research design and tools for online research," *The SAGE Handbook of Online Research Methods* , N. G. Fielding, R. M. Lee, and G. Blank, Eds., 55 City Road, London: SAGE Publications Ltd, 2017, pp. 57-75. [Online]. Available: https://methods.sagepub.com/book/the-sage-handbook-of-online-research-methods-second-edition

[31] V. Toepoel, "Online survey design," *The SAGE Handbook of Online Research Methods* , N. G. Fielding, R. M. Lee, and G. Blank, Eds., 55 City Road, London: SAGE Publications Ltd, 2017, pp. 184-202. [Online]. Available: https://methods.sagepub.com/book/the-sage-handbook-of-online-research-methods-second-edition

[32] F. Corea, "AI knowledge map: How to classify AI technologies," in *An introduction to data. Studies in big data* , vol. 10: Springer, 2019.

[33] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, "Operationalising AI ethics: Barriers, enablers and next steps," *AI & SOCIETY,* 2021, doi: 10.1007/s00146-021-01308-8.

[34] K. Calagna, B. Cassidy, and A. Park, "Enterprise risk management: Realize the full potential of artificial intelligence." [Online]. Available: https://www.coso.org/Documents/Realize-the-Full-Potential-of-Artificial-Intelligence.pdf

[35] C. Canca, "Operationalizing AI ethics principles," *Communications of the ACM,* vol. 63, no. 12, pp. 18-21, 2020, doi: 10.1145/3430368.

[36] K. Jia and N. Zhang, "Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines," *Electronic Markets,* 2021, doi: 10.1007/s12525-021-00480-5.

[37] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.,* vol. 1, no. 9, pp. 389-399, 2019/09/01 2019, doi: 10.1038/s42256-019-0088-2.

[38] National Institute of Standards and Technology, "NIST draft - Taxonomy of AI risk." [Online]. Available: https://www.nist.gov/system/files/documents/2021/10/15/taxonomy_AI_risks.pdf

[39] C. Verbano and K. Venturini, "Development paths of risk management: approaches, methods and fields of application," *Journal of Risk Research,* vol. 14, no. 5, pp. 519-550, 2011, doi: 10.1080/13669877.2010.541562.

[40] L. Floridi *et al.* , "AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines,* vol. 28, no. 4, pp. 689-707, 2018, doi: 10.1007/s11023-018-9482-5.

[41] A. Hamedi, "The concept of justice In greek philosophy (Plato and Aristotle)," *Mediterranean Journal of Social Sciences,* vol. 5, no. 27 P2, p. 1163, 12/10 2014. [Online]. Available: http://www.richtmann.org/journal/index.php/mjss/article/view/5193.

[42] M. Velasquez, C. Andre, T. Shanks, and M. J. Meyer, "Thinking ethically," *Issues in Ethics,(August),* pp. 2-5, 2015.

[43] J. G. Dahl, M. P. Mandell, and M. E. Barton, "Ethical frameworks of "tomorrow's business leaders"," *International Journal of Value Based Management,* vol. 1, no. 2, pp. 65-81, 1988, doi: 10.1007/BF03184883.

[44] J. Bethem, G. Frigo, S. Biswas, D. C. Tyler, and M. Pasqualetti, "Energy decisions within an applied ethics framework: an analysis of five recent controversies," (in English), *Energy, Sustainability and Society,* vol. 10, no. 1, Dec 2020 2020, doi: 10.1186/s13705-020-00261-6.

[45] N. Asgary, A. Walle, and S. P. Saraswat, "Ethical foundations and managerial challenges: The strategic implications of moral standards," (in English), *Journal of Leadership, Accountability and Ethics,* vol. 11, no. 2, pp. 89-98, Jun 2014 2014.

[46] United States Census Bureau. *2019 SUSB annual data tables by establishment industry. Datasets: U.S. & states, NAICS, detailed employment sizes (U.S., 6-digit and states, NAICS sectors), and U.S., NAICS sectors, larger employment sizes up to 20,000+* . [Online]. Available: https://www.census.gov/data/tables/2019/econ/susb/2019-susb-annual.html

[47] K. A. Robinson, I. J. Saldanha, and N. A. McKoy, "Development of a framework to identify research gaps from systematic reviews," *Journal of Clinical Epidemiology,* vol. 64, no. 12, pp. 1325-1330, 2011/12/01/

2011, doi: https://doi.org/10.1016/j.jclinepi.2011.06.009.

[48] S. Lawn, C. Oster, B. Riley, D. Smith, M. Baigent, and M. Rahamathulla, "A Literature Review and Gap Analysis of Emerging Technologies and New Trends in Gambling," *International Journal of Environmental Research and Public Health,* vol. 17, no. 3, p. 744, 2020, doi: 10.3390/ijerph17030744.

[49] Organisation for Economic Co-operation and Development, "Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems," *OECD Digital Economy Papers* , no. 312*,* doi: 10.1787/008232ec-en.

[50] D. Peters, K. Vold, D. Robinson, and R. A. Calvo, "Responsible AI—Two frameworks for ethical design practice," *IEEE Transactions on Technology and Society,* vol. 1, no. 1, pp. 34-47, 2020, doi: 10.1109/TTS.2020.2974991.

[51] V. Vakkuri and K.-K. Kemell, "Implementing AI ethics in practice: An empirical evaluation of the RESOLVEDD strategy," presented at the International Conference on Software Business (ICSOB), Cham, 2019, Paper Presentation. [Online]. Available: https://dx.doi.org/10.1007/978-3-030-33742-1_21.

[52] K. Forbes, "Opening the path to ethics in artificial intelligence," *AI and Ethics,* no. 1, pp. 297-300, 2021, doi: 10.1007/s43681-020-00031-2.

[53] I. Georgieva, C. Lazo, T. Timan, and A. F. van Veenstra, "From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience," *AI and Ethics,* vol. 2, no. 4, pp. 697-711, 2022/11/01 2022, doi: 10.1007/s43681-021-00127-3.

[54] L. Munn, "The uselessness of AI ethics," *AI and Ethics,*2022, doi: 10.1007/s43681-022-00209-w.

[55] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, "Explaining the principles to practices gap in AI," *IEEE Technology and Society Magazine,* vol. 40, no. 2, pp. 81-94, 2021, doi: 10.1109/mts.2021.3056286.

[56] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance," *AI and Ethics,* vol. 2, no. 4, pp. 603-609, 2022/11/01 2022, doi: 10.1007/s43681-022-00143-x.

[57] B. Shneiderman, *Human-centered AI* . Oxford University Press, 2022.

[58] D. J. Teece, "Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance,"*Strategic Management Journal,* vol. 28, no. 13, pp. 1319-1350, 2007, doi: 10.1002/smj.640.

[59] A. Nair, E. Rustambekov, M. McShane, and S. Fainshmidt, "Enterprise risk management as a dynamic capability: A test of its effectiveness during a crisis," *Managerial and Decision Economics,* vol. 35, no. 8, pp. 555-566, 2014, doi: 10.1002/mde.2641.

[60] Y. Bogodistov and V. Wohlgemuth, "Enterprise risk management: a capability-based perspective," *The Journal of Risk Finance,* vol. 18, no. 3, pp. 234-251, 2017, doi: 10.1108/jrf-10-2016-0131.

[61] A. Nishimura, "Comprehensive opportunity and lost opportunity: Control model and enterprise risk management," *International Journal of Business and Management,* vol. 10, no. 8, pp. 73-87, 2015, doi: 10.5539/ijbm.v10n8p73.

[62] A. Shore, "Talking about facial recognition technology: How framing and context influence privacy concerns and support for prohibitive policy," *Telematics and Informatics,* vol. 70, pp. 1-12, 2022/05/01/ 2022, Art no. 101815, doi: 10.1016/j.tele.2022.101815.

[63] S. Zhang, Y. Feng, and N. Sadeh, "Facial recognition: Understanding privacy concerns and attitudes across increasingly diverse deployment scenarios," presented at the Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021), Virtual Conference, 2021, Paper Presentation. [Online]. Available: https://www.usenix.org/conference/soups2021/presentation/zhang-shikun.

[64] S. Du and C. Xie, "Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities," *Journal of Business Research,* vol. 129, pp. 961-967, 01/01 2020, doi: 10.1016/j.jbusres.2020.08.024.

[65] International Organisation for Standardization, "Risk management - ISO 31000 brochure." [Online]. Available: https://www.iso.org/publication/PUB100426.html

[66] National Institute of Standards and Technology, "AI risk management framework concept paper." [Online]. Available: https://www.nist.gov/system/files/documents/2021/12/14/AI%20RMF%20Concept%20Paper_-13Dec2021_posted.pdf

[67] P. R. Nagbøl, O. Müller, and O. Krancher, "Designing a risk assessment tool for artificial intelligence systems," presented at the The Next Wave of Sociotechnical Design. DESRIST., 2021. [Online]. Available: https://dx.doi.org/10.1007/978-3-030-82405-1_32.

[68] T. Madiega, "Briefing: Artificial Intelligence Act," pp. 1-12. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

[69] M. Kop, "EU artificial intelligence act: The European approach to AI," *Transatlantic Antitrust and IPR Developments,* 2021. [Online]. Available: https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/.

[70] European Union. "Shaping Europe's digital future." https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed 9 August 2022.

[71] AIAAIC. "Zoom AI emotion recognition." https://www.aiaaic.org/aiaaic-repository/ai-and-algorithmic-incidents-and-controversies/zoom-ai-emotion-recognition (accessed 23 August 2022.

[72] R. Lutz, "Incident number 95," *Artificial intelligence incident database* , S. McGregor, Ed.: Responsible AI Collaborative, 2019. [Online]. Available: https://incidentdatabase.ai/cite/95#r1397

[73] M. C. Tremblay, A. R. Hevner, and D. J. Berndt, "Focus groups for artifact refinement and evaluation in design research," *Communications of the association for information systems,* vol. 26, no. 1, pp. 599-618, 2010, Art no. 27, doi: 10.17705/1CAIS.02627.

[74] P. Eriksson and A. Kovalainen, *Qualitative methods in business research: A practical guide to social research* , 2nd. ed. Sage, 2016.

[75] J. Saldana, *The coding manual for qualitative researchers* , 3rd ed. London: Sage, 2016.

[76] AIID. "AI incident database." https://incidentdatabase.ai/ (accessed 3 August 2022.

[77] S. McGregor, "Preventing repeated real world AI failures by cataloging incidents: The AI incident database," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 35, no. 17, pp. 15458-15463, 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17817.

[78] AIAAIC. "AIAAIC - AI, algorithmic and automation incidents & controversies." https://www.aiaaic.org/home (accessed 3 August 2022.

[79] AI Global. "Where in the world is AI?" https://map.ai-global.org/ (accessed 3 August 2022.

[80] C. Cichy and S. Rass, "An overview of data quality frameworks," *IEEE Access,* vol. 7, pp. 24634-24648, 2019, doi: 10.1109/access.2019.2899751.

[81] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems,* vol. 12, no. 4, pp. 5-33, 1996, doi: 10.1080/07421222.1996.11518099.

[82] E. Awad, S. Dsouza, A. Shariff, I. Rahwan, and J.-F. Bonnefon, "Universals and variations in moral decisions made in 42 countries by 70,000 participants," *Proceedings of the National Academy of Sciences,* vol. 117, no. 5, pp. 2332-2337, 2020, doi: 10.1073/pnas.1911517117.

[83] R. Inglehart and W. E. Baker, "Modernization, cultural change, and the persistence of traditional values," *American Sociological Review,* vol. 65, no. 1, pp. 19-51, 2000, doi: 10.2307/2657288.

[84] R. Dale, "GPT-3: What's it good for?," *Natural Language Engineering,* vol. 27, no. 1, pp. 113-118, 2021, doi: 10.1017/s1351324920000601.

[85] Y. K. Wong, "Practicality of Quantum Computing & AI," *International Journal of Engineering Trends and Applications,*vol. 9, no. 2, pp. 6-11, 2022, doi: 10.33144/23939516/IJETA-V9I2P2.

[86] M. Kop, "Establishing a legal-ethical framework for quantum technology," *Yale Law School, Yale Journal of Law & Technology (YJoLT), The Record,* 2021. [Online]. Available: https://yjolt.org/blog/establishing-legal-ethical-framework-quantum-technology.