An Enterprise Risk Management Framework to Design Pro-Ethical AI Solutions

Quintin McGrath[1] ( ) [0000-0003-1616-7668], Alan Hevner[2] [0000-0003-4953-3900] and Gert-Jan de Vreede[3] [0000-0001-6909-9836]

Muma College of Business, University of South Florida, Tampa, FL, USA
[1] qmcgrath@usf.edu, [2] ahevner@usf.edu, [3] gdevreede@usf.edu

Abstract. The wide use of Artificial Intelligence (AI) has immediate business benefits for an organization and its stakeholders through efficiency gains, greater repeatability, and newdebue6 1 (a)1b0 0 6 349e (a)1b0 0 6 349 m0.03 Twohr d ( )Tj -0.0046

being terminated by an algorithm with little human intervention or chance for appeal has a negative impact on the business and its stakeholders

Ethics is the lens through which the rightness and wrongness of decisions and business practices are evaluated. So, for long-term benefit, the development of AI System Solutions (AISS) should be guided by ethical thinking to ensure the greatest positive benefit and the least possible harm to the business and its stakeholders [3]. Deviating from what is generally considered right results in an ethical risk. This balancing act is common for business leaders who must assess opportunities and their associated risks using Enterprise Risk Management (ERM) frameworks [4, 5]. ERMs maximize business value by assessing the probability of the occurrence of a positive or negative event and managing its anticipated impacts. For the business leader, AISS are complex, adding new risks and opportunities. This is not only because of AI complexity (e.g., self-learning potential, intelligent capabilities, and inscrutability) but also because of its emergent behaviors and tighter integration with human users in a socio-technical system. Because of this complexity and the dynamic, emergent behavior of AISS, we contend that a more nuanced ERM framework with an ethical focus is sorely needed.

This research in progress paper proposes the dynamic management of related AI Ethical Risks to enhance an organization's risk posture. An Active Design Science Research (DSR) approach will be used with the research objective to build an enhanced Enterprise Risk Management (ERM) which has the goal of maximizing busi-

business value, which we contend, includes value to all stakeholders, by considering business opportunities and their associated risks. Steinberg et al. defines ERM as follows:

> Enterprise risk management is a process, effected by an entity s board of direc-

AI Ethical Risks are the unexpected negative consequences of design, development, and operational actions

Fig. 1. The e-ERM Framework

3.1 Phase 1: Identify

The first phase defines identifies the business risks and opportunities related to the AISS and its environment. The focus is therefore on the socio-technical aspects of the AISS, the envelope within which it operates, and the AI Ethical Principles that will be used to guide the identification of the AI related ethical risks.

- The AISS as a Socio-Technical System. Asatiani et al. [24] point out that, because of the growth in AI's capabilities, focus has shifted from considering only the technical aspects of an AISS to incorporating the social component. The proposed e-ERM framework sees an AISS as a socio-technical system composed of (a) machine aspects (e.g., the algorithm and data), (b) human actors (e.g., the business goals and decisions associated with the system), and (c) the interaction between the machine, the human, and the associated controls that integrate the two.
- Envelopment. Because of the breadth of the risks and opportunities of AISS there is a need to establish operational boundaries. Nagbøl et al. [25] recommend a bounding approach called envelopment, based on Robbins, [26] in their AI risk assessment tool. In this way, defining what an AISS may and may not do. We use this same envelopment approach to constrain the socio-technical AISS and limit its scope to be within a defined set of the business goals, system objectives, and stakeholders.
- AI Ethical Principles. This study uses a broad perspective on ethics, considering both not doing the wrong things but also encouraging right actions. When the use of AI is

trustworthiness. In a similar way we apply the bioethical principles as a foundation across the AISS lifecycle. We define human aspects as the ethical principles and risk of the human part of the socio-technical system, like business goals and fairness. The machine aspects are the ethical principles and risks implemented in the algorithms and data of an AISS. The interaction aspects focus on the risks relating to the AI Ethical Principles of accuracy and reliability of the solution, its transparency, and explainability.

3.2 Phase 2: Assess

As part of the assessment phase of the ERM framework, a risk assessment tool is planned. This tool identifies, analyzes, and evaluates risks and opportunities to aid decision-making relating to the socio-technical AISS. An example of such a tool is Nagbøl et al.'s [25] Artificial Intelligence Risk Assessment (AIRA) tool. The AIRA consists of three modules that target specific groups to allow for an efficient risk assessment. The first module considers business needs for the AISS. The second focuses on the technical details of the system. The final module is a synthesis of the outputs of the previous modules balancing the AISS's business and technical aspects.

The outputs from our assessment phase are a clearly defined AISS system, with its context, AI capabilities, and objectives, along with a prioritized list of risks and opportunities that should be acted upon in the next phase. Most important here are the ethical risks associated with the three elements of the socio-technical AISS. This assessment phase provides a static or point-in-time assessment of the planned AISS and is based on known risks. Because of the changing nature of the AISS and its environment, continual reassessment will be needed (see below.)

3.3 Phase 3: Act

With the output from the previous phase, we can enter the next phase where the appropriate action for each of the risks can be determined, e.g., avoid, mitigate, share, transfer, or accept the risks [4]. To enable focused action, we will create a Risk Reference Database (RRD). This is based on the analysis of publicly available AI incident databases combined with input on what are considered best practices from subject matter experts. This RRD uses the output from the assessment phase, including AI capabilities, the AISS context and its goals, and the prioritized ethical risks to identify

This research will therefore focus on organizations developing AISS for these parts of their HR processes.

To drive the dynamic nature of the E-ERM, the feasibility of three approaches to continuous monitoring and sensing will be tested. The first is a manually triggered model to link the latest AI incidents to AI Ethical Principles and to store these in the RRD. The second is applying AI machine learning to the publicly available information on incidents to regularly enrich the RRD. Finally, the use of a Generative Adversarial Networks (GAN) approach to actively track the ethical risk profile of the SAIS compared to the emerging best practices in the RRD will be evaluated. We propose the following stages for our future research directions:

- Problem Space/Diagnosis Stage: We have completed the first iteration of the DSR process through the creation of the E-ERM framework artifact. We will validate the various parts of the model, its relevance, and potential usefulness using expert interviews and focus groups of subject matter experts. We will use the results of the focus group to refine the model.
- Solution Space/Design Stage: We will then move to the solution space and enter the design stage via the following paths: (1) conduct interviews with around 20 IT Leaders to analyze successful strategies to link an organization's AI Ethical Principles to risk mitigating practices; (2) in order to incorporate the end user and indirect stakeholder's perspective, analyze the various existing AI incident databases to establish a relationship between incidents and the relevant AI Ethical Principles and the related inc(in)-7 (c)EMC /L

elicited from successful organizations, allowing for effective continuous improvement of the AISS and proactively addressing potential ethical risks. Further contributions include the synthesis of the work by NIST, ISO, and others and defining the AI principles supporting responsible, ethical AISS. Our development approach supports an important set of HR applications for study.

In terms of limitations, as this is early work, there is much that still needs to be tested and validated. Also, future research will investigate approach for the inclusion of input from those indirectly impacted by the ERM and the resulting AISS.

References

1. Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J.R., Tchatchouang

icopm (S)7.8ciTmdini(pr)9(.)1.6 (o (S)7.8)1.6 (obl Tc 0 e)16.2 ()1.6 (oap4.2 (h f)73.7 (or)1.7 (t)18.)15 (dBa

15. Francis, R.D.: Ethical Risk Management. In: Farazmand, A. (ed.) Global Encyclopedia of Public Administration, Public Policy, and Governance, pp. 1-5. Springer International Publishing (2016)
16. Ruehle, C.R.: Investigating Market and Regulatory Forces Shaping Artificial Intelligence Adoptions. Muma Business Review 1, 177-192 (2020)
17. Sheth, A., Yip, H.Y., Iyengar, A., Tepper, P.: Cognitive Services and Intelligent Chatbots: Current Perspectives and Special Issue Introduction. IEEE Internet Computing 23, 6-12 (2019)
18. Marshall, T.E., Lambert, S.L.: Cloud-Based Intelligent Accounting Applications: Accounting Task Automation Using IBM Watson Cognitive Computing. Journal of