Ethical Risk Assessment of AI in Practice Methodology: Process-oriented Lessons Learnt from the Initial Phase of Collaborative Development with Public and Private Organisations in Norway

Natalia Murashova
Department of Teacher Education
Østfold University College
Halden, Norway
Email: natalia.murashova@hiof.no

Diana (Saplacan) Lindblom
Department of Informatics
University of Oslo
Oslo, Norway
Email: dianasa@ifi.uio.no

Aida Omerovic
SINTEF Digital
Oslo, Norway
Email: aida.omerovic@sintef.no

Heidi E. I. Dahl
Posten Bring
Oslo, Norway
Email: heidi.dahl@posten.no

Leonora Onarheim Bergsjø
Østfold AI Hub
Østfold University College
Oslo, Norway
Email: leonora.bergsjo@hiof.no

***Abstract*—Artificial Intelligence (AI) and its ethical implications are not new for academia and business. Challenges of embedding principles for ethical AI in practice are obvious and even though the gap between theory and practice is decreasing, it does not meet the urgent need for responsible technology development and deployment. Embedding ethical principles in existing risk assessment practices is a novel, process-oriented approach that can contribute to operationalising AI ethics in organisational practice. This paper elaborates on initial phase of collaborative development of ethical risk assessment of AI methodology, involving private and public organisations in Norway. We reflect upon our experience and present key takeaways in a form of three lessons learnt from embedding a Model-based security risk analysis method (CORAS) and a Story Dialog Method (SDM) in the initial phase of the collaborative methodology development. This study concludes that ethical risk assessment of AI in practice is feasible and explores design issues related to cross-sectoral settings, flexibility of the methodology, and power-relationship.**

Keywords-AI ethics in practice, ethical risk assessment, cross-sectoral collaboration, methodology development.

I. INTRODUCTION

Current transformations of the work environments due to wide introduction of AI systems call for new approaches to assess and manage benefits and risks created by these systems. AI, understood as an umbrella term in computer science for different computing techniques enabling machines to mimic "complex human skills", holds many promises and hopes together with uncertainty and fears [1][2]. Since ungoverned AI comes together with social and environmental implications, organisations that design and deploy AI are obliged to identify and mitigate possible risks throughout design process and application lifecycle.

To pursue ethical design and deployment of AI systems, a principled approach is commonly used to guide the process of development and deployment of AI systems [3][4]. Taking its popularity and relative accessibility for both practitioners and researchers, it has become a leading approach to ethical AI resulting in over 100 ethical guidelines in private and public sectors [5]-[7]. Empirical studies showed that the impact of ethical guidelines on ethical decision-making of the professionals is very low [8]-[10]. This suggests a gap between ethical AI principles available and their relevance for organisational practice. The challenge lies not only in contextualizing the principles for each stage of development of AI systems or a use case, but also in what professional competences needed to practice, promote and deploy ethical AI [11][12]. It is argued that solely ethical principles in place are not enough to responsibly navigate a complex landscape of AI applications in organisations and that embedding a "risk-oriented multi-stakeholder approach" in the assessment and management procedures is a key to effective governance of AI [13]-[16].

Prior research suggests that, to achieve ethical AI in practice, we as research community, society, governments, businesses, have to create a common language, provide equal opportunity for stakeholder involvement and create a system of incentives for ethical decision-making [6][17][18]. This study gave an opportunity to public and private organisations in Norway to take an active part in shaping ethical risk assessment of AI in practice methodology alongside with researchers from different disciplines, such as ethics, data science, risk analysis, pedagogy and social sciences. This study investigates the following research question; What are the key lessons learnt from the initial phase of ethical risk assessment methodology development? How these lessons can be applied in the next phase of the methodology development?

Through analysis of the initial stage of the ethical risk assessment of AI methodology development, this paper contributes to the developing body of knowledge on ethical AI in practice and elaborate on the outcomes of the collaborative processes involving academic, industrial and public sector partners. In addition, we reflect upon the lessons learnt to guide the methodology calibration further.

The rest of this paper is organized as follows. Section II describes the background followed by the related work section.

Section IV introduces CORAS and SDM as a theoretical base for this study. Section V focuses on the methods including approaches implemented during data collection, analysis and participant recruitment together with ethical consideration. Section VI presents the findings in a form of the learnt lessons. Section VII dives into discussion connected with the results together with limitations. The acknowledgement, conclusion and future work suggestions close the article.

II. Background

Rapidly emerging AI technology poses many challenges to social and organisational structures worldwide including environmental costs, social implications and ethical dilemmas [19]. Norway, among the other European countries, is on the regulatory side, adopting European Union's AI Act, but at the same time has an ambition to build a national infrastructure for AI in the public sector by 2030 [20][21]. Therefore, this invites public and private organizations to use AI systems for digitalisation and innovation [21][22].

Norwegian organisations that are developing and deploying AI systems are under pressure since they must comply with local and international data protection regulations, ethical norms, established sectorial traditions, and most importantly innovate at the same time. While legal regulations and sectoral traditions can be addressed through existing mechanisms, ethical AI is new and not well-established concept in some sectors, thus, the public and private organizations stand in front of new challenges of implementing ethical AI in practice [4][23].

In regard to the above-mentioned points, Ethical risk assessment of AI in practice (ENACT) project has a goal of creating such methodology that can benefit Norwegian businesses to evaluate and mitigate risks connected with design and deployment of AI in their organization using "ethical lens". Among the other objectives is tailoring this methodology to be adaptive, scalable, process-oriented and applicable to different organisational contexts and AI applications [24].

The ENACT consortium comprises industrial partners from public and private sectors, including medical services, finance, logistics, welfare service and education, in addition to researchers from SINTEF (lead), Østfold University College (ØUC), the University of Oslo (UiO), and the Norwegian University of Science and Technology (NTNU). ENACT is funded by the Research Council of Norway (2023-2027).

III. Related work

A. Principles for ethical AI in practice

Several governing bodies, individuals and research initiatives have released guidelines to promote responsible AI and ensure its development and deployment in an ethical manner [3][18][25]-[28]. In the last 5 years, a principled approach became the most favourable one in the literature [5][17]. Originally adopted from the bioethics and medical research, it is centred around respect of autonomy, non-maleficence, beneficence and justice [29]. Floridi has extended above-mentioned principles by adding "explicability or transparency" as a new enabling principle [25]. Policymakers, organisations, philosophers, AI researchers and practitioners have contextualized these principles resulting in over 100 public and private guidelines across the globe [5].

Despite a high level of abstraction, many attempts have been made to operationalise the ethical principles within different industrial domains [10][30][31]. Even though the organizations acknowledge importance of AI ethics and aim to be proactive, the range of employed strategies, that work in practice, seem limited, in comparison with the wide range proposed in existing literature [10][17][32]. However the gap between principle and practice is decreasing, but many organisations struggle to use ethical AI frameworks due to the ambiguity of ethical principles and variety of approaches to its design in practice [6][33].

A principled approach to AI ethics faced some criticism from being ineffective in practice to lacking a long-standing professional tradition [4][8][34]. The challenge lies not only in contextualizing the principles for each stage of development of AI systems or a use case but also in what professional competences are needed to practice, promote and deploy ethical and responsible AI [11][12].

B. Ethical risk assessment of AI in practice

Operationalising AI ethics is a challenging and multifaceted task that usually involves various stakeholder groups and processional competences. There is always an ethical aspect in a risk that is affected by our moral views and values [23][35]. Ethical Risk Assessment for identification and mitigation of possible impacts is usually performed by the ethical committees and boards and has not been standardised in regard to AI systems [36]-[38]. Due to the rapid expansion of AI use and design in organisations, the ethical dimension of risk has become a prominent topic in the discussion of social and environmental impacts of AI. Previous studies indicate that one of the challenges of ethical risk assessment of AI lies in qualitative nature of the assessment that, compared to the ordinary risk assessment approaches, is not quantifiable and rarely effective via box checking [14][36]. In addition, single sector studies showed that ethical discussions are not traditionally embedded in some sectors, compared to healthcare for example, which makes ethical risk assessment difficult and highly abstract [4][38].

Several studies have addressed the questions of how ethical risk assessment can be performed in organisations. For example, Tartaro et al. attempted to embed an ethical dimension in the risk assessment procedures through a four staged process including open ended questions, risk grouping, Likert scale evaluation (numerical value assignment), and risk identification and visualisation [14]. This study confirmed the limitations of the check-list approach and binary questions (yes/no) for complex ethical risk assessment. The study concluded that the open-ended nature of questions to prompt the participants contributes to inclusive discussion and increases the support of the risk mitigation measures by the stakeholders [14].

Felländer et al. have employed a multi-disciplinary approach to achieve data-driven risk assessment for ethical AI [39]. Using expert knowledge to build the definitions and establish the requirements for cross-sectoral application, authors highlighted the difficulty of creating ethical risk assessment tool applicable to different sectors and relatable to practical, real-world problems [39].

To address power imbalances and enhance a complex analysis of normative issues under risk assessment process, Krijger explored a relational approach by proposing a triad of decision-maker, risk-exposed and beneficiary to understand and qualitatively analyse how risk is distributed and aligned with political and social moral of the stakeholders [40].

IV. Theoretical framework

To interpret, translate and integrate ethical norms, frameworks and guidelines into organisational practice, the limitations of the principled approach can be addressed through a process-oriented perspective through using, for example, CORAS and SDM, which are described further.

A. Model-based method for security risk analysis

Compared to traditional risk analysis, which are based on "failure-oriented" aspects of a system, model-based risk assessment includes different aspects of the system, giving a holistic view on the risks connected with it [41].

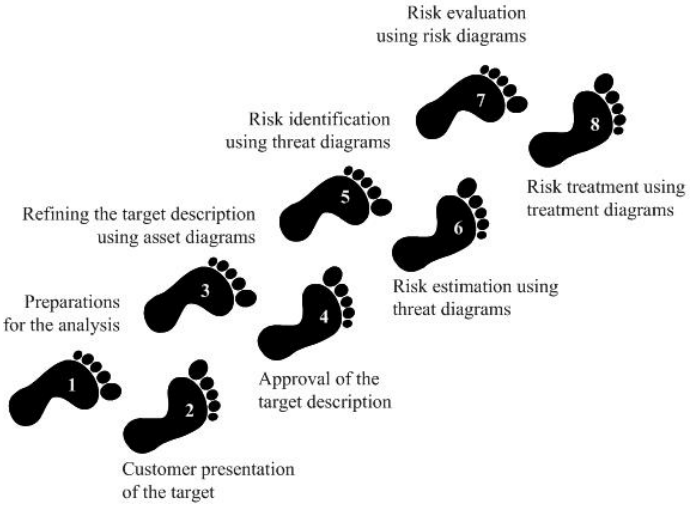

Figure 1. CORAS steps for conducting security risk analysis [42].

Model-based method for security risk assessment, also known as CORAS, is conducted in three phases: context establishment, risk assessment and risk treatment including sub-processes which contain a specific set of steps presented in Figure 1 [42].

The context establishment phase specifies the target and the scope of the analysis. During the risk assessment phase, the relevant risks are identified, evaluated and estimated. The stakeholders (organisation representatives) are being engaged on different steps of the process, allowing a variety of input and expertise shape the risk assessment outputs. The last phase focuses on treatment of the evaluated risks. Among the strengths of this approach is graphical style of the communication, visual modelling, constructive use of language and tighter integration of the assessment outputs in the system development processes [43]. As for the limitations, the practitioners noted that the CORAS language can be perceived as "too simplistic and cumbersome to use" [44].

Since this approach has been designed for defensive risk analysis of the assets with a focus on security, some fields might find it difficult to steer the discussion using such terms as "threat scenario" or "vulnerability". In addition, the risk analysis is performed by an external team which in practice is expensive and difficult to scale up.

B. Making sense of organisational experiences through Story Dialog Method

SDM is both a data collection and a data analysis method [45][46]. SDM has its roots within critical pedagogy, constructivism and feminism, and critical social sciences [45][46]. It is based on a structured dialogue and on participants' stories.

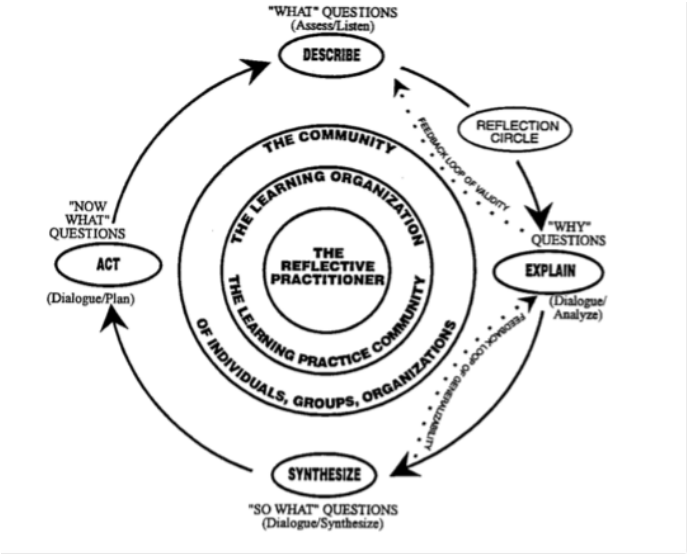

Figure 2. Story Dialogue Method [46].

A story can be described as a "self-interview" in a particular situation. Each participant has a dedicated role: storyteller, story listener, or story recorder, along with facilitators and observer roles. SDM also has values as its point of departure, which makes it relevant for the ENACT methodology. Further, the method is based on a structured dialogue, following four stages and questions related to each of the stages (see also Figure 2):

- Describe (where WHAT-type of questions are asked)
- Explain (WHY-type of questions are asked)
- Synthesis (where SO WHAT-type of questions are asked)
- Action (NOW WHAT-type of actions are asked)

While the method was initially intended to be used by a homogeneous group of participants (e.g., participants from the

same organization) in physical settings, the method lately has been used and adapted in a variety of settings with participants coming from various organizations and backgrounds, as well as online [47]-[51].

One of the benefits of this method, especially in power uneven collaborative environments, is active engagement of the participants into knowledge co-creation. It contributes to promotion of equality and inclusion in the dialog about ethical AI in organizations.

V. METHOD

In attempt to address existing limitations and challenges, ENACT methodology is collaboratively developed, tested, evaluated and validated by an interdisciplinary consortium of researchers and small and large organisations from the Norwegian private and public sector. Development of ENACT methodology is grounded in collaboration and iterative adjustment of the methods presented in the previous section.

A. Study context

The initial phase of ENACT methodology development and testing took place in September – December 2024. The working group held over 25 meetings in total (internal and with the stakeholders) to discuss the adjustments of the CORAS and integration of the SDM methods, to adopt it to digital workshop format and to accommodate the sectoral needs of the partners.

To develop a methodology for ethical risk assessment of AI which could work in practice, the working group attended to several issues. Together, the working group examined the challenges and opportunities for harmonizing and synthesising major steps of the CORAS with ethical core issues in practical settings. In addition, the analysis context was narrowed down to one use-case that all participating organisations had in common. Participating organisations have collectively agreed on using Microsoft Co-Pilot transcription tool as a use-case for ethical risk assessment.

Another important decision was made in regard to the participants and is described further.

B. Participants in a transdisciplinary collaboration

First, to cover multiple perspectives of ethical AI, the working group was comprised of the researchers with expertise in ethics, risk management, technology and pedagogy. Secondly, to attend to practical issues of business settings, the work group adopted a transdisciplinary approach. Representatives from businesses and organisations were included in the design of the ENACT methodology, and a representative from one of the businesses was included in the working group that designed the initial testing described in this paper.

Participant selection for the workshops was purposefully delegated to the ENACT-consortium business partner representatives who were our main contact persons throughout the process. These allowed us to speed up the trust bonding process with the workshop participants through the contact person who already had workplace connections with the employees [52]. To include different organisational perspectives in the collaborative process, the working group agreed on recruiting participants from different sectors and disciplines to create interdisciplinary and cross-sectoral participant pool for all three workshops.

In total, we have involved 14 unique participants in all the workshops and 29 in total across all workshops (non-unique). The workshop participants had various levels of seniority in the organisation and years of experience in their positions. Therefore, it was important to address invisible hierarchy in the group dynamic in the methodology design and try to "level up the field players" through compatible facilitation techniques and embedding SDM [53].

C. Data collection

Due to multifaceted nature of the collaborative process and the relationships that are formed, this study has comprised different data sources including: meeting notes, observational notes from the workshops, pre- and post-workshop survey, PowerPoint slides from the workshops [54]. Table I shows an overview over the data collection process.

Two facilitators and two to three observers from the working group were allocated per workshop. The facilitators were responsible for introducing the theme of the workshop, the concepts and facilitating the discussion. The observers had a passive role during the workshops but at the same time took process-oriented notes in addition to participants' reflections and ideas presented.

Together with the participants, it was decided to manually collect observational notes for securing open dialog and freedom of expression under the workshops. Several participants had made it clear before the workshop that they preferred manual data collection methods rather than audio/video recordings of the workshops, which would have affected their behavoiur. It was agreed that audio or video recording of the workshops will affect the behaviour and nature of the interactions among the participants.

Observational notes were structured according to the workshop slides (pre-selected themes) and business partner involved in the discussion (concerns, comments, ideas). In addition, we collected observers' and facilitators' post-workshop notes about the process. All the notes were gathered in a separate file after the workshops and then used for methodology refinement and adjustment of the workshop content.

D. Analysis

Using a combination of explorative approaches, such as iterative assessment of the notes, collective reflection and synthesis, helped make sense of the processes occurring in the collaborative environments [55]. Since most of the data were gathered from the participant interactions (surveys, workshops, observation notes), workshop PowerPoints and reflections of the working group, a combination of analysis techniques was applied, as follows.

TABLE I
OVERVIEW OF THE PROCESS

Workshops and CORAS steps	Workshop 1: Establishment of the context	Workshop 2: Risk assessment	Workshop 3: Risk assessment
Time and format	60 mins, Digital workshop	60 mins, Digital workshop	60 mins, Digital workshop
Structure of the workshop	1. Introduction, information and expectations 2. Key features of use case, stakeholders, timeframe 3. Values 4. High level analysis	1. Information and expectations 2. Values – identifying values and discussion 3. Scenarios – identifying values and discussion 4. Summary	1. Information and expectations 2. Scenarios 3. Group brainstorming 4. Summary of group discussion and initial measures 5. Summary and evaluation
Participants	10 participants 2 facilitators 3 observers	10 participants 2 facilitators 3 observers	9 participants 2 facilitators 2 observers
Ground for methodology adjustment	Pre-workshop survey, Work group meetings	Post workshop survey, Work group meeting, ENACT business partner meeting, ENACT project meeting	Work group meeting, ENACT business partner meeting
Concepts	Actors	Values	Scenarios
Documentation	Pre-workshop survey, 8.5 pages of structured notes	Post-workshop survey, 8 pages of structured notes	2 pages of notes

The main goal of the analysis process was to interpret textual data and researchers' observations form the workshops to explore potential lessons for ethical risk assessment of AI methodology adjustment and tuning. This study focused on a systematic process analysis that was inspired by the thematic analysis [56].

Texts that were systematically produced by the observers under the workshops and the discussions taking place right after were structured according to each workshop. These were then read multiple times by several of members of the working group, to preliminary map the process and aspects that needed adjustment. In addition, post-workshop working group meetings were held to share the experience after the workshop and reflect on the process together.

After reading the notes multiple times and summing up the reflections we have come up with analytical notes to guide the methodology adjustment and identify learning points to be taken to the next phase of testing. Sensemaking of qualitative data for this study emerged from multiple levels of analysis including participant interaction between each other, different sectors, different seniority levels. The process that we were trying to understand was unfolding in a continuum of the collaborative process rather than in a hierarchically structured manner [57]. That is why the findings are presented in a form of lessons, overarching the major take-aways from the workshops.

E. Ethical considerations

All parts of this study were conducted in line with national and international guidelines for research ethics and research integrity [58]-[60]. To attend to the rights, interests and well-being of the participants, best practices of consent, privacy and data protection were deployed. The study was planned in accordance with ENACT Data Management Plan and reported to the Norwegian Agency for Shared Services in Education and Research. All gathered data assets were stored at the Services for Sensitive Data at the University of Oslo.

In addition to protecting personal data of the participants, we had to protect organizational data too. It was therefore communicated to the participants that all the confidential information shared with us under the workshops will not be included in the overall findings and results to preserve organizational confidentiality. Moreover, the study was conducted with participants representing different organizational entities and this added another layer of complexity for ethical considerations in the research process. To create a safe environment for sharing relevant info, and to protect business representatives from sharing protected information, the researchers had several dialogues with representatives from the business partners concerning which parts of the methodology they felt safe to test in a cross-sectorial group both before and between the workshops.

VI. FINDINGS

In the course of collaborative methodology development, we have adjusted several aspects of the process to address organisational needs of the participants. The analysis of these adjustments resulted in three process-oriented lessons that are presented in this section.

A. Lesson 1. The scope of ethical risk in cross-sectoral settings

Since sectoral traditions are different, a common "analysis context" has to be identified and addressed in the methodology design to facilitate the process of risk identification and assessment. But even when the common ground is found, not all participants are ready to discuss risk treatment practices openly with other organisational sectors present. On the one hand, the difference in sector specific use-cases provided a broader scope of the discussion by incorporating different perspectives. Additionally, it helped participants to centre their reflections around ethical perspectives of risk identification and analysis and gave structure to the discussion, contributing to bonding and blending of participants' cross-sectoral experience. On the other hand, the cross-sectoral settings of the workshops created some boundaries for engagement in a deeper discussion of

the scenario-based assessment because of resistance from the participants that occurred due to different sectoral traditions and business confidentiality.

B. Lesson 2. Flexible methodology helps to address organisational needs

Our analysis suggests that ENACT methodology must accommodate organisational needs and sectoral demands in addition to addressing changing organisational dynamics and competitive AI landscape [61]. Flexibility can be achieved through feedback loops and iterative content adjustment. Feedback loops as part of the tailoring help to guide the process toward current organisational demands and make the workshops relevant for all the participants. The partner organisations wished for a methodology that can be easily embedded in everyday practice and will be resource efficient. We have observed that digital format and selected timeframe of the workshops worked satisfactory. In addition, a desired depth of the discussion was not reached despite all the adjustment efforts made by the working group. Introduction of the new concepts, used to guide the discussion, took away the time from ethical reflections and discussion.

C. Lesson 3. Easing power-relationship for structured dialog and critical reflection

Balanced and equal engagement of the participants from different sectoral traditions and seniority level was a difficult task. Our observations concluded that to promote equal power relations, the number of participants per workshop (in plenum) should be reduced to actively include everyone in the discussion. Using elements from the SDM (facilitating empowerment through giving the opportunity to all stakeholders to voice their opinions and worries) helped us to even the field players through structured yet opened environment of engagement. This approach contributed to creations of the safe space where all the participants, regardless of their position in organisation, could share their "stories" and contribute to the process. It was challenging to engage every single one of the participants at the same time, but we managed to give the opportunity to everyone to engage in the discussion and share their worries and views.

In some cases, the participants should be separated in smaller groups depending on the goal of the discussion. For fostering cross-sectoral reflections and a broader scope it is useful to separate the participant from their fellow colleagues. This fosters cross-sectoral reflections and broadens the scope, in addition to easing out existing hierarchical structures among the participants from the same organisation. On the other hand, it was admitted by the participants that they had to "hold back" some information in group and plenum discussions due to a business confidentiality.

VII. Discussion

The initial phase of collaborative development of the ENACT methodology was a demanding and rewarding process at the same time. Structural adjustments of the methodology have been enhanced by reflections of the working group and feedback loops from the participants. In the process of tuning the working group have identified following challenges that should be addressed for further development of ethical risk assessment of AI in practice including

- sectoral tradition (e.g., similarities and differences between the domain of ethics and security standards with respect to risk assessment)
- group dynamics (e.g., power dynamics in the group, business integrity, approaches to elicit organisational needs)
- confidential information of organisational practices
- format of the workshops, which had to be realistic (e.g., time, digital or physical meetings, resources required) if the businesses were to use the methodology in real world settings

When the overall workshop structure is adaptive, it can suit different professional competences, use cases and value landscapes. Two CORAS steps, including establishment of the context and risk assessment, serve as base for the ethical risk assessment and can be aligned with already established procedures of risk assessment in the organisations. Due to cross-sectoral nature of collaboration, not all participants were ready to discuss risk treatment in depth, explaining it by business confidentiality and difference in sectoral traditions. In previous studies, cross sectoral nature of the risk assessment was admitted being challenging, due to different sectoral traditions and value misalignment [38][39].

According to our results, choosing one use-case or scenario-based approach does benefit the process in term of direction to the discussion about the application of AI systems and stakeholders involved in its design and use. But at the same time, practice showed that it was difficult for facilitators to elicit concrete scenario-based solutions for ethical dilemmas that were identified in the first two steps of the assessment.

Participation and participant selection is about power, therefore the participants that are chosen to "sit at the table" matter in terms of diversity and inclusion of ethical risk assessment of AI [31]. While at this stage, we did not integrate all the SDM aspects, we adopted a few central elements from SDM: the idea of a structured dialogue, and we ensured that all participants gave their input on each question. Embedding SDM form the very beginning would, in theory, contribute to evenly distributing time for each participant to join the discussion and express their ideas during the workshop. This would even out the dynamics in uneven power environment.

In respect to the format, digital workshops worked well in terms of maximizing the number and variety of participants. As for the drawbacks, digital format created a mediated space for the interactions which had complicated the communication and the flow of the process [62]. Because of the time constrains, not all participants had the opportunity to present their ideas, in addition to that, some of the ongoing discussions had to be interrupted due to the time constrains which negatively affected depth and quality of the reflection processes.

A. Limitations

This study involved participants from a small sample of Norwegian public and private organizations. This implies that organizational culture in these sectors can differ from the international context and other local contexts, and so could the organisational needs related to AI development and deployment. Moreover, the purposeful sampling of the participants might have resulted in a biased representation of employees. In addition, the test design excluded non-Norwegian speakers and employees not involved in AI system development and use, which might be relevant stakeholders. The test also included several businesses, and to protect their interests, collaborative efforts resulted in a small sample of themes and methods agreed upon for testing.

In addition, Microsoft Co-Pilot was chosen as a common AI use-case for the cross-sectoral discussion and analysis, which limits the scope of this study to this use scenario. Observational notes taken under the workshop were taken by different researchers resulting in possible biases in the reported observations.

VIII. Conclusion and future work

Increasing presence of AI systems in everyday work of organisations signifies a need for profound changes in the way we understand, identify and assess the risks connected to its use and design. The quantity and quality of engagement with AI ethics in organisation determine the depth of the discussion and the pool of risks identified. Despite well-articulated principles for ethical AI, practice shows that their practical use in organisations is ambiguous and limited.

Our study presented several methodological reflections about ethical risk assessment of AI in practice methodology development based on Norwegian organisational context. In the course of this study, we have managed to test different workshop structures and facilitation tools based on the organisational needs that were elicited through regular feedback loops, observations and working group discussions. In a course of collaborative design, we elicited process-oriented benefits and challenges in a form of three lessons focusing on the scope of ethical risk assessment in cross-sectoral setting, flexibility of the methodology and the power relationship occurring under the workshops.

This testing phase has resulted in preliminary skeleton of ENACT methodology that needs future adjustments. Our findings suggest a need for further research on ENACT methodology implementation for example in single sector settings, different sized organisational structures and national contexts. This can enrich the understanding of the pitfalls, sector-specific value landscapes and needs that are crucial for efficient ethical risk assessment of AI in practice.

In addition, there is a need for expansion of the philosophical discourse around moral values involved in AI risk assessment and what role ethics plays in it [35].

Acknowledgment

The research project Ethical Risk Assessment of Artificial Intelligence in Practice (ENACT) is funded by the Research Council of Norway (project number 338170). The ENACT consortium comprises industrial partners from the public and private sectors, including medical services, finance, logistics, welfare service and education, in addition to researchers from SINTEF, Østfold University College (ØUC), the University of Oslo (UiO), the Norwegian University of Science and Technology (NTNU) and NORA.ai (Norwegian AI Research Consortium).

References

[1] H. Sheikh, C. Prins, and E. Schrijvers, "Artificial intelligence: definition and background," Mission AI. Research for Policy, pp. 15-41, Jan 2023, doi.org/10.1007/978-3-031-21448-6_2.

[2] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?". Proc. ACM Conference on Fairness, Accountability, and Transparency, Mar 2021, pp. 610-623, doi.org/10.1145/3442188.3445922.

[3] High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*. [Online]. Available from https://shorturl.at/2XleY 2025.04.10.

[4] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," Nature Machine Intelligence, vol. 1, no. 11, pp. 501-507, Nov 2019, doi.org/10.1038/s42256-019-0114-4.

[5] T. Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," Minds and Machines, vol. 30, no. 1, pp. 99-120, Feb 2020, doi.org/10.1007/s11023-020-09517-8.

[6] E. Prem, "From ethical AI frameworks to tools: a review of approaches," AI and Ethics, vol. 3, pp. 699–716, Feb 2023, doi.org/10.1007/s43681-023-00258-9.

[7] D. Schiff, J. Borenstein, J. Biddle, and K. Laas, "AI ethics in the public, private, and NGO sectors: A review of a global document collection," IEEE Transactions on Technology and Society, vol. 2, no. 1, pp. 31-42, Mar 2021, doi:10.1109/TTS.2021.3052127.

[8] A. McNamara, J. Smith, and E. Murphy-Hill, "Does ACM's code of ethics change ethical decision making in software development?," Proc. 26th ACM joint meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Oct 2018, pp. 729-733, doi.org/10.1145/3236024.3264833.

[9] L. Munn, "The uselessness of AI ethics," AI and Ethics, vol. 3, no. 3, pp. 869-877, Aug 2023, doi.org/10.1007/s43681-022-00209-w.

[10] J. C. Ibáñez and M. V. Olmeda, "Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study," AI & Society, vol. 37, no. 4, pp. 1663-1687, Dec 2022, doi.org/10.1007/s00146-021-01267-0

[11] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," Proc. CHI Conference on Human Factors in Computing Systems, Apr 2020, pp. 1-16, doi.org/10.1145/3313831.3376727.

[12] S. Strauß, ""Don't let me be misunderstood" Critical AI literacy for the constructive use of AI technology," Journal for Technology Assessment in Theory and Practice, vol. 30, no. 3, pp. 44-49, Dec 2021, doi.org/10.14512/tatup.30.3.44

[13] R. Clarke, "Principles and business processes for responsible AI," Computer Law & Security Review, vol. 35, no. 4, pp. 410-422, Aug 2019, doi.org/10.1016/j.clsr.2019.04.007.

[14] A. Tartaro, E. Panai, and M. Z. Cocchiaro, "AI risk assessment using ethical dimensions," AI and Ethics, pp. 1-8, Jan 2024, doi.org/10.1007/s43681-023-00401-6

[15] E. Vyhmeister, G. Castane, P.-O. Östberg, and S. Thevenin, "A responsible AI framework: pipeline contextualisation," AI and Ethics, vol. 3, no. 1, pp. 175-197, Apr 2023, doi.org/10.1007/s43681-022-00154-8.

[16] B. W. Wirtz, J. C. Weyerer, and I. Kehl, "Governance of artificial intelligence: A risk and guideline-based integrative framework," Government Information Quarterly, vol. 39, no. 4, p. 101685, Oct 2022, ISSN 0740-624X, doi.org/10.1016/j.giq.2022.101685.

[17] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices," Science and Engineering Ethics, vol. 26, no. 4, pp. 2141-2168, Aug 2020, doi.org/10.1007/s11948-019-00165-5.

[18] V. Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Cham: Springer International Publishing AG, pp. 93-105, 2019.
[19] UN Human Rights Office of the High Commissioner, *Taxonomy of Human Rights Risks Connected to Generative AI*, [Online] Available from https://shorturl.at/KXSB4 2025.04.09.
[20] European Union, *The EU Artificial Intelligence Act*. [Online]. Available from: https://artificialintelligenceact.eu/ 2025.04.05.
[21] Norwegian Ministry of Local Government and Modernisation *National Strategy for Artificial Intelligence*. [Online]. Available from: https://shorturl.at/4oiLr 2025.04.09.
[22] K. Bjørgo. *Utnytte mulighetene i kunstig intelligens. Exploiting the opportunities of artificial intelligence*. [Online]. Available from: https://shorturl.at/Tg5ES 2025.04.09.
[23] S. O. Hansson, "How to perform an ethical risk analysis (eRA)," Risk Analysis, vol. 38, no. 9, pp. 1820-1829, Sep 2018, doi/10.1111/risa.12978
[24] ENACT. "About us". [Online]. Available from: https://www.enactai.no/about 2025.04.02.
[25] L. Floridi et al., "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations," Minds and Machines, vol. 28, pp. 689-707, Dec 2018, doi.org/10.1007/s11023-018-9482-5.
[26] UNESCO, *Recommendation on the ethics of artificial intelligence*. [Online]. Available from: https://shorturl.at/60xL7.
[27] OECD, *Recommendation of the council on artificial intelligence* OECD/LEGAL/0449.
[28] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82-115, Jun 2020, doi.org/10.1016/j.inffus.2019.12.012.
[29] T. Beauchamp and J. Childress, "Principles of biomedical ethics: marking its fortieth anniversary (1979)," The American Journal of Bioethics, vol. 19, pp. 9-12, Oct 2019, doi.org/10.1080/15265161.2019.1665402.
[30] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Putting AI ethics into practice: The hourglass model of organizational AI governance," ArXiv preprint, Jan 2022, arXiv:2206.00335.
[31] J. Ayling and A. Chapman, "Putting AI ethics to work: are the tools fit for purpose?," AI and Ethics, vol. 2, no. 3, pp. 405-429, Aug 2022, doi.org/10.1007/s43681-021-00084-x.
[32] B. C. Stahl, J. Antoniou, M. Ryan, K. Macnish, and T. Jiya, "Organisational responses to the ethical issues of artificial intelligence," AI & Soc, vol. 37, no. 1, pp. 23-37, Mar 2022, doi.org/10.1007/s00146-021-01148-6.
[33] J. Morley et al., "Operationalising AI ethics: barriers, enablers and next steps," AI & Soc, pp. 1-13, Feb 2023, doi.org/10.1007/s00146-021-01308-8.
[34] T. Hagendorff, "AI virtues. The missing link in putting AI ethics into practice," ArXiv preprint, Feb 2020, arXiv:2011.12750.
[35] S. O. Hansson, "Risk and ethics: Three approaches," in Arguing about Science: Routledge, 2012, pp. 629-640.
[36] A. F. Winfield and K. Winkle, "RoboTed: a case study in Ethical Risk Assessment," arXiv preprint, Sep 2020, arXiv:2007.15864.
[37] R. D. Bernabe et al., "The risk-benefit task of research ethics committees: An evaluation of current approaches and the need to incorporate decision studies methods," BMC Med. Ethics, vol. 13, pp. 1-9, Apr 2012, doi.org/10.1186/1472-6939-13-6.
[38] N. Chugh, "Risk assessment tools on trial: Lessons learned for "Ethical AI" in the criminal justice system," 2021 IEEE International Symposium on Technology and Society (ISTAS), IEEE, 2021, pp. 1-5, doi:10.1109/ISTAS52410.2021.9629143.
[39] A. Felländer, J. Rebane, S. Larsson, M. Wiggberg, and F. Heintz, "Achieving a data-driven risk assessment methodology for ethical AI," DISO, vol. 1, no. 2, p. 13, Aug 2022, doi.org/10.1007/s44206-022-00016-0.
[40] J. Krijger, "What About Justice and Power Imbalances? A Relational Approach to Ethical Risk Assessments for AI," DISO, vol. 3, no. 3, p. 56, Oct 2024,doi.org/10.1007/s44206-024-00139-6.
[41] B. A. Gran, R. Fredriksen, and A. P. J. Thunem, "An Approach for Model-Based Risk Assessment," Lecture Notes in Comput. Sci., vol 3219, 2004, doi.org/10.1007/978-3-540-30138-7_26. In International Conference on Computer Safety, Reliability, and Security, pp. 311-324. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
[42] M. S. Lund, B. Solhaug, and K. Stølen, Model-driven risk analysis: the CORAS approach. Springer, pp. 23-43, 2010.
[43] R. Fredriksen et al., "The CORAS framework for a model-based risk management process," Proc. 21st International Conference (SAFECOMP), Sep 2002, pp. 94-105, ISBN 978-3-540-45732-9.
[44] K. Stølen, "The CORAS method: process, concepts and notation". SINTEF. [Online]. Available from: https://shorturl.at/XTifB 2025.04.07.
[45] R. Labonté, "Reflections on stories and a story/dialogue method in health research," International Journal of Social Research Methodology, vol. 14, no. 2, pp. 153-163, Aug 2011, doi.org/10.1080/13645579.2010.492131.
[46] R. Labonte, J. Feather, and M. Hills, "A story/dialogue method for health promotion knowledge development and evaluation," Health Education Research, vol. 14, no. 1, pp. 39-50, Feb 1999, doi.org/10.1093/her/14.1.39.
[47] D. Saplacan, J. Herstad, M. N. Elsrud, and Z. Pajalic, "Reflections on using Story-Dialogue Method in a workshop with interaction design students," Proc. CEUR Workshop (AVI 18), May 2018, pp.34-43, URI https://hdl.handle.net/10642/7422.
[48] D. Saplacan, J. Herstad, A. Mørch, A. Kluge, and Z. Pajalic, "Inclusion through design and use of digital learning environments: issues, methods and stories," Proc. 10th Nordic Conference on Human-Computer Interaction, Sep 2018, pp. 956-959, doi.org/10.1145/3240167.3240264.
[49] H. W. Moen, L. E. Kjekshus, D. Saplacan, and Z. Pajalic, "How do professionals in home-based health care services react to a new trust-based way of working?–A story dialogue study," Social Sciences & Humanities Open, vol.11, p. 101251, Dec 2024, doi.org/10.1016/j.ssaho.2024.101251.
[50] D. Saplacan, T. Schulz, J. Torresen, and Z. Pajalic, "Health Professionals' Views on the Use of Social Robots with Vulnerable Users: A Scenario-Based Qualitative Study Using Story Dialogue Method," The 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), IEEE, Aug 2023, pp. 421-428.
[51] Z. Pajalic, D. Saplacan, I. Borgen, S. E. G. Olsen, and N. N. Wesseltoft-Rao, "Female university academics' reflections on the development of their academic careers in the Norwegian higher education context," Social Sciences & Humanities Open, vol. 8, no. 1, p. 100548, May 2023, doi.org/10.1016/j.ssaho.2023.100548.
[52] A. A. Kliskey et al., "Building trust, building futures: Knowledge co-production as relationship, design, and process in transdisciplinary science," Frontiers in Environmental Science, vol. 11, p. 1007105, Feb 2023, doi.org/10.3389/fenvs.2023.1007105.
[53] L. Wood and M. McAteer, "Levelling the playing fields in PAR: The intricacies of power, privilege, and participation in a university–community–school partnership," Adult Education Quarterly, vol. 67, no. 4, pp. 251-265, Nov 2017, doi.org/10.1177/0741713617706541.
[54] S. Davenport, J. Davies, and C. Grimes, "Collaborative research programmes: building trust from difference," Technovation, vol. 19, no. 1, pp. 31-40, Nov 1998, doi.org/10.1016/S0166-4972(98)00083-2.
[55] G. Guest, K. M. MacQueen, and E. E. Namey, Applied Thematic Analysis. SAGE Publications, Inc., pp. 3-20, 2012.
[56] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101, Jul 2008, doi.org/10.1191/1478088706qp063oa.
[57] A. Langley, "Strategies for theorizing from process data," Academy of Management Review, vol. 24, no. 4, pp. 691-710, Oct 1999, doi.org/10.2307/259349.
[58] E. Staksrud et al., "Guidelines for Research Ethics in the Social Sciences and the Humanities". [Online]. Available from: https://shorturl.at/inpI4 2025.04.10.
[59] The Norwegian National Committee for Research Ethics in Science and Technology, "Guidelines for research ethics in science and technology". [Online]. Available from: https://shorturl.at/k13CY 2025.04.10.
[60] P. J. Drenth, "A European code of conduct for research integrity," [Online]. Available form: https://shorturl.at/1KtpI 2025.04.10.
[61] T. H. Davenport and R. Ronanki, "Artificial intelligence for the real world," [Online]. Available from: https://shorturl.at/COinT 2025.04.10.
[62] A. Beaulieu and A. Estalella, "Rethinking research ethics for mediated settings," Information, Communication & Society, vol. 15, no. 1, pp. 23-42, Jun 2011, doi:10.1080/1369118X.2010.535838.