Navigating Responsible AI Adoption

Daniela Oliveira[a]
Independent Researcher, Canada

Keywords: Responsible Artificial Intelligence, Responsible Adoption, Artificial Intelligence Governance, Change Management, Knowledge Management, Organizational Learning.

Abstract: Responsible Artificial Intelligence has been a largely discussed topic among organizations that develop or are aiming to regulate Artificial Intelligence (AI) solutions. Much less attention has been given to organizations willing to adopt AI in a responsible manner. Organizations that do not develop AI need practical guidance on how to implement Responsible AI principles. This contribution outlines the challenges organizations face integrating the Responsible AI paradigm and suggests some solutions.

1 INTRODUCTION

One of the most distinguishing characteristics of Artificial Intelligence (AI) technologies from other kinds of technological solutions is how intrinsically connected to data these technologies are. This characteristic brings a different dimension to software development: customizing an AI solution to a specific context is highly dependent on this context (Friedler et al., 2021).

As a consequence, the impact and risks of an out-of-the-lab solution can be extremely different from the impact and risks of the implemented solution. The fact that AI technologies mostly automate cognitive processes solidifies this condition. Adopting an AI solution is not only about training workforce to use a new technological solution, but also reaching a point where the workforce thinking is enhanced or replicated in a satisfactory manner - and not diminished - by the technological solution. A point found in the middle ground between adapting the AI solution to the context of application and integrating humans to the development process. The development and adoption processes of the technological solution are then rather close, sometimes even overlapping, when it comes to AI, compared to traditional technologies.

AI solutions have additional challenges: they need a conscious effort to predict and reduce the harm they can cause and become what is known as a Responsible AI solution (Celdran et al., 2023; Siala & Wang, 2022; Université de Montréal, 2018). The development or adoption processes have also to consider the possible harms the AI solutions can provoke.

Considering the harm an AI solution can provoke has its share of context-specific considerations, as each organization has its own culture and processes, influenced by national and regional culture, market and regulatory practices, to mention only some of the aspects that compose the success or failure of organizations. AI adoption can impact much more than the technology infrastructure or the data management practices of an organization. It can motivate changes in human resources practices (Tursunbayeva & Renkema, 2022), brand, reputation (World Economic Forum, n.d.) and knowledge management (Jarrahi et al., 2022), to name a few, and present issues throughout the lifecycle of a product or service. AI solutions can motivate situations where unexpected human behavior in processes where AI solutions are integrated produce unexpected outcomes, possibly leading to physical harm; they can exacerbate gaps in workplace training; motivate unclear identification of human and AI's work, with mismatching expectations and evaluation practices; monitoring mechanisms that are centered in AI solutions and disregard their interaction with employees and other humans; policies that were not updated to reflect the complexity of collaborating with AI solutions; unaligned human / AI quality assurance initiatives; and non-optimal timeliness.

[a] https://orcid.org/0000-0001-9285-0173

Oliveira, D.
Navigating Responsible AI Adoption.
DOI: 10.5220/0012262600003598
In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - Volume 2: KEOD*, pages 339-345
ISBN: 978-989-758-671-2; ISSN: 2184-3228

AI solutions have not only to be customized in terms of what functionalities they should address, but also how these functionalities should be addressed by the solution. The remaining portrait is that no two AI implemented solutions are alike in terms of impacts and risks and should be even less in their adoption processes.

Adoption processes are led by the organization willing to successfully integrate a technological solution in their operations or strategic activities. In AI solutions, they can represent more than half of the analytics budget (Fountaine et al., 2019). However, little attention has been granted to AI adoption processes focusing on Responsible AI principles, making it difficult for organizations to plan resources in order to responsibly adopt an AI solution.

This paper aims to shed light on adoption processes, and the number and order of challenges involved in these processes. It is organized as follows: this Introduction, an overview of the Responsible AI paradigm in section 2 and of challenges concerning the adoption of AI solutions in section 3; a definition of the Responsible AI Adoption process in section 4; an overview of the Responsible AI Adoption process challenges in section 5, followed by the conclusion of the contribution.

2 RESPONSIBLE AI

Responsible AI is a term often associated with a global movement to ensure AI solutions' risks are addressed and mitigated. Responsible AI technology has been defined as fair and accountable (Agarwal & Mishra, 2021; Siala & Wang, 2022), explainable (Agarwal & Mishra, 2021), respecting privacy and fostering transparency (Siala & Wang, 2022; B. K. Vassileva, 2021), trustworthiness and empathy (Siala & Wang, 2022). Responsible AI guidelines such as the Montreal Declaration for a Responsible Development of AI (Université de Montréal, 2018) and the Responsible Microsoft Standard (Microsoft Corporation, 2022) have been created by government bodies, AI development companies and research institutes, to name a few. Lukkien et al. (2023) recognize a "growing prevalence of frameworks, principles, and guidelines to inform responsible AI innovation" (p.156) but deplore that most of them present high-level principles with excessive room for interpretation along with limited practical measures in specific context of use.

3 AI ADOPTION

AI has been the object of unprecedent technological enthusiasm, to the fact that organizations are willing to make drastic changes in favour of its adoption. Alignment between business and information technology objectives was one of the top concerns two decades ago (Reich & Benbasat, 2000), but it has been argued by an AI development company that "companies must break down organizational and cultural barriers that stand in AI's way" (Fountaine et al., 2019). In practice, how much change should an organization absorb in favour of AI adoption to remain competitive? This is a question highly bounded to the organizational context.

The AI adoption process must be adapted to reflect an organization's regional position, business domain or the challenges of a specific kind of AI technology. The AI adoption process might range from stablishing key performance indicators that are meaningful to the organization to fostering cultural changes, passing by scrutinizing technical and business interoperability potentially affected by the AI solution adoption.

The AI adoption process must also be adapted to reflect the needs surrounding organizational culture regarding technology. Humans may trust technological solutions more than they should, a phenomenon known as cognitive complacency (Jarrahi, 2019), even when their outputs seem to be wrong or inadequate, a situation of automation bias (Skitka et al., 2000). This factor may be more prominent in some organizations than in others (Alon-Barkat & Busuioc, 2023). In addition, great appetite for AI may lead to overly confident attitudes (Perry et al., 2022) that make for longer turnaround reactions. Without organizational culture changes, the collaboration between humans and AI may generate deceiving results.

In addition, AI solutions implementation must be preceded by a risk assessment analysis (Brand, 2022; Cebulla et al., 2022; Clarke, 2019; Leijnen et al., 2020; Nagbøl et al., 2021; Oliveira & Dalkir, 2022; Qiang et al., 2023). Risks may be inherent to the solution, or a consequence of the application of the solution in the use context. Concerning population characteristics, an example of inherent risk of a solution is one that discriminates a portion of the population (Mattu, 2016), while one that is a consequence of the application in the use context is the difference between the population characteristics in the time range used for training the model and when the model was put to use (Suresh & Guttag, 2021). Risks inherent to solutions can be analysed by developers and academic actors, but risks that stem from the application of the solution in the use context depend on an analysis of the use context against the documentation of the AI solution following Responsible AI principles (Mitchell et al., 2019), which may vary from one organization to another.

After the implementation of an AI solution, monitoring the output of the solution and engagement levels among the direct and indirect user population helps ensure the predicted return on investment is realized without loss of client base or the addition of excessive restrictions to operations. The exercise of identifying what to monitor, however, has to begin before an AI solution is implemented.

4 RESPONSIBLE AI ADOPTION

Much attention has been given to Responsible AI solutions development which, arguably, must integrate different kinds of stakeholders, according to practitioners, academicians and vendors alike (Minkkinen et al., 2023; Obermeyer et al., 2021).

However, organizations that do not develop AI systems seem to have been excluded from the discussion. An analysis of Responsible AI studies and frameworks yielded only three types of stakeholders: individuals and national or international bodies and organizations involved with AI regulation: technology companies, professional bodies or research institutes (Deshpande & Sharp, 2022). This situation creates a gap between Responsible AI in theory and in practice. While AI can be the focus of activity of academicians and vendors, it is not the case for most organizations willing to adopt AI. These organizations may have less knowledge of AI than developers might assume (Richards et al., 2020) but are, nevertheless, an important element in the AI environment. They are the organizations holding the data used to train AI models for specific tasks, the organizations offering AI-based services to individuals and other organizations, the organizations that allocate financial and human resources for the acquisition of AI systems and the first organizations to be subject to negative operational, reputational, relational, and legal impacts of AI if they were to take place. These organizations need clear substantiation as to how Responsible AI principles can be flexibly attuned in context (Bærøe et al., 2020; Lukkien et al., 2023). Guiding principles can evolve an organization's thinking on Responsible AI, "but they are not sufficient for implementing responsible AI principles across everything from development to acquisition to operations" (Probasco, 2022, p. 1).

Due to the multifaceted impact of AI, much of the Responsible AI gains can only be achieved if followed by a Responsible AI Adoption process. In order to operationalize the adoption of AI solutions respecting responsible principles, Leijnen et al. (2020) invoke the importance of assessing AI solutions before implementation, and the inclusion of usability principles and agile approaches. Adopting AI in a responsible manner means that practical aspects of the adoption were analysed and accounted for in the decision to adopt the AI solution. It may mean as little organizational reflection as ensuring the solution is used as per recommended by its developers (Mitchell et al., 2019) or as much as triggering business processes; data visualisation initiatives; stakeholders participation; strategic positioning and information technology architecture analysis and evaluation, among other possibilities.

Some of the simple aspects of Responsible AI Adoption are related to communication and feedback: removing obstacles for people to voice concerns over a specific solution. However, organizations need to reflect on who, apart the user population, should have the ability to voice concerns over a specific solution. Employees and managers may contribute in decisive ways not otherwise considered (Rolls Royce, 2021) and change management, particularly participative, and effective integration of domain knowledge have been correlated with successful AI adoption (von Richthofen et al., 2022). Organizations also need to reflect on how to conduct this input collection and treatment.

Some of the more complex Responsible AI Adoption aspects are related to the organization's positioning in the AI environment. For example, how valuable are the organization's data to AI providers? With this understanding, some organizations negotiated reciprocal agreements that consider the value of the data involved (Siala & Wang, 2022).

How far have potential providers of an AI solution adopted Responsible AI principles? Responsible AI Adoption may help envision responsible initiatives to palliate shortcomings of an AI solution. For example, where the replication of bias present in the training data can be an issue (Au Yeung et al., 2023), the Responsible AI Adoption process can include retraining the solution with data that does not contain harmful biases and a proof-of-concept decisive stage.

The Responsible AI Adoption can be a differential for organizations enforcing a growth of their AI solutions portfolio aligned with Responsible AI values. It can also place the organization a step forward in regions where regulations are being developed to hold organizations adopting AI accountable of the due diligence implied in the Responsible AI paradigm.

5 RESPONSIBLE AI ADOPTION CHALLENGES

5.1 Selecting Responsible AI Frameworks and Principles

A Responsible AI Adoption strategy should reflect the principles of Responsible AI applicable and contain actionable measures in the context of use. While the number of Responsible AI guidelines containing principles has been growing, the number of guidelines allowing for operationalization of Responsible AI are still scarce (Lehoux et al., 2023; Lukkien et al., 2023; Narayanan & Schoeberl, 2023).

In a study that analyses the crossroads between long-term care, AI and responsible innovation, Lukkien et al. (2023) argue that Responsible AI can only be fostered with practical measures that apply the principles conveyed. The study aimed to identify concrete measures to influence design and/or implementation of actual AI solutions in a specific context of use. The kind of guidance the authors sought were named “process-based frameworks” by Narayanan and Schoeberl (2023): frameworks offering a blueprint helping organizations to prioritize aspects of system design, identify accountability lines, establish the infrastructure, resources and capabilities needed to operationalize Responsible AI. These frameworks fall under the category of "operational tools" (Lehoux et al., 2023), documents that provide in-depth hands-on guidance on a particular issue with detailed explanations, real-word examples, step-by-step activities and further resources.

A portrait of the challenge faced by organizations in order to benefit from Responsible AI Adoption is reflected in the research effort of the study of Lukkien et al. (2023): from 3,339 documents advocating for Responsible AI, only 8 contained practical measures in the context of use. Current awareness of Responsible AI guidelines can be time-consuming.

Process-based guidelines are more sensitive to the context of use than Principles-based guidelines. In 2021, there were at least 170 frameworks or tools to support Responsible AI operationalization (Deshpande & Sharp, 2022). The number of process-based guidelines is growing, but the challenge to select and apply these tools persist (Lehoux et al., 2023; Narayanan & Schoeberl, 2023). The work of Narayanan and Schoeberl (2023) eases the Responsible AI Adoption process in an interesting way. The authors created a taxonomy to help organizations navigate process-based Responsible AI frameworks based on the study of 45 generic Responsible AI frameworks. The Matrix for Selecting Responsible AI Frameworks was created to assist organizations identifying frameworks that meet their specific needs.

However, in contexts where process-based frameworks are not available, the need to evaluate impact and mitigate risks are still present. An ethics framework, guiding the interpretation and translation of Responsible AI principles into actionable measures is therefore helpful. Principles-based frameworks have great variance in scope (Fjeld et al., 2020; Jobin et al., 2019; Lehoux et al., 2023), even though some consensus can be verified (Morley et al., 2020). The context of use of the AI solution might also be the object of more than one principles-based guideline. An ethics framework can guide the organization in the interpretation and translation of more than one principles-based guideline.

5.2 Finding the Right Ethics Framework

Applying ethics to business has always been a delicate endeavour (Murray, 1997) and it is no different when it comes to AI. An ethics framework that is aligned with the organizational culture can help identify risks, prioritize initiatives, obtain buy-in, guide training and communication around the adoption of AI and help the interpretation of principles-based guidelines and their translation into actionable measures. Ethics framework do not address AI challenges themselves. Instead, they offer ways to approach AI challenges and solutions.

For instance, Verbeek and Tijink (2020) argue for a three-step approach: 1) considering the technology in context; 2) involving actors, values and effects; 3) identifying options for action, which can be: a) co-creation with users and b) ethics by design, in context and in use.

The co-creation approach is also preconized by Bruneault et al. (2022). This framework specifies individual, organizational and social attitudes and structure. All the actors involved, the authors argue, should continuously question pre-conceived ideas, should employ the nuancing that characterizes concrete applications and should recognize knowledge limitations at any given moment.

5.3 Informing AI Governance

Effective AI governance should shape reality according to governing concepts and also empower agents of this reality to shape their governing concepts (Noiseau, 2023). This two-way interdependence can be used as a way to avoid effective accountability (Floridi, 2019), but coupled with a Responsible AI Adoption process, motivated by actual applications and aligned with the organizational culture, can be an effective path to actually implement and maintain Responsible AI solutions. The measurable performance indicators and clear criteria for monitoring risks resulting from the Responsible AI Adoption process can contribute for a more effective AI governance.

6 CONCLUSION

To be effective, the Responsible AI paradigm demands guidelines that are both broad in coverage and specific in advice, as well as the identification of impact and risks of the adoption of AI solutions in the domain and context of use. Whereas articulating the Responsible AI paradigm is a task better performed by academia and regulatory bodies, the identification of actionable measures with little room for interpretation demands specific domain and organizational knowledge.

This contribution outlined activities related to the development and implementation of AI in organizations in order to foster a responsible, continuous, incremental and aligned AI adoption. These activities are impacted by the Responsible AI paradigm but aim actionable measures, adapted to the domain and context of use of the AI solution. This contribution suggests the need for talent, time, budget and particular expertise to promote Responsible AI Adoption processes.

REFERENCES

Agarwal, S., & Mishra, S. (2021). *Responsible AI: Implementing ethical and unbiased algorithms*. Springer.

Alon-Barkat, S., & Busuioc, M. (2023). Human-AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice. *Journal of Public Administration Research and Theory*, *33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Au Yeung, J., Kraljevic, Z., Luintel, A., Balston, A., Idowu, E., Dobson, R. J., & Teo, J. T. (2023). AI chatbots not yet ready for clinical use. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1161098

Bærøe, K., Miyata-Sturm, A., & Henden, E. (2020). How to achieve trustworthy artificial intelligence for health. *Bulletin of the World Health Organization*, *98*(4), 257–262. https://doi.org/10.2471/BLT.19.237289

Brand, D. J. (2022). Responsible Artificial Intelligence in Government: Development of a Legal Framework for South Africa. *eJournal of eDemocracy and Open Government*, *14*(1), 130–150. Scopus. https://doi.org/10.29379/jedem.v14i1.678

Bruneault, F., Laflamme, A. S., & Mondoux, A. (2022). *Former à l'éthique de l'IA en enseignement supérieur: Référentiel de compétence*. SocArXiv. https://doi.org/10.31235/osf.io/38tfv

Cebulla, A., Szpak, Z., Howell, C., Knight, G., & Hussain, S. (2022). Applying ethics to AI in the workplace: The design of a scorecard for Australian workplace health and safety. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01460-9

Celdran, A. H., Kreischer, J., Demirci, M., Leupp, J., Sanchez, P. M., Franco, M. F., Bovet, G., Perez, G. M., & Stiller, B. (2023). *A Framework Quantifying Trustworthiness of Supervised Machine and Deep Learning Models*. 2938–2948.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law and Security Review*, *35*(4), 410–422. Scopus. https://doi.org/10.1016/j.clsr.2019.04.007

Deshpande, A., & Sharp, H. (2022). Responsible AI Systems: Who are the Stakeholders? *AIES 2022 - Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 227–236. Scopus. https://doi.org/10.1145/3514094.3534187

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (SSRN Scholarly Paper 3518482). https://doi.org/10.2139/ssrn.3518482

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, *32*(2), 185–193. https://doi.org/10.1007/s13347-019-00354-x

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization: Technology isn't the biggest challenge. Culture is. *Harvard Business Review*.

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2021). The (Im)possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, *64*(4), 136–143. https://doi.org/10.1145/3433949

Jarrahi, M. H. (2019). In the age of the smart artificial intelligence: AI's dual capacities for automating and informating work. *Business Information Review*, *36*(4), 178–187. Scopus. https://doi.org/10.1177/0266382119883999

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2022). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*. https://doi.org/10.1016/j.bushor.2022.03.002

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: The global landscape of ethics guidelines. *arXiv.Org*. https://doi.org/10.48550/arXiv.1906.11668

Lehoux, P., Rivard, L., de Oliveira, R. R., Mörch, C. M., & Alami, H. (2023). Tools to foster responsibility in digital solutions that operate with or without artificial intelligence: A scoping review for health and innovation policymakers. *International Journal of Medical Informatics*, *170*, 104933. https://doi.org/10.1016/j.ijmedinf.2022.104933

Leijnen, S., Aldewereld, H., van Belkom, R., Bijvank, R., & Ossewaarde, R. (2020). An agile framework for trustworthy AI. *NeHuAI@ ECAI*, 75–78.

Lukkien, D. R. M., Nap, H. H., Buimer, H. P., Peine, A., Boon, W. P. C., Ket, J. C. F., Minkman, M. M. N., & Moors, E. H. M. (2023). Toward Responsible Artificial Intelligence in Long-Term Care: A Scoping Review on Practical Approaches. *Gerontologist*, *63*(1), 155–168. Scopus. https://doi.org/10.1093/geront/gnab180

Mattu, J. A., Jeff Larson,Lauren Kirchner,Surya. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Microsoft Corporation. (2022). *Microsoft Responsible AI Standard*. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2023). Co-Shaping an Ecosystem for Responsible AI: Five Types of Expectation Work in Response to a Technological Frame. *Information Systems Frontiers*, *25*(1), 103–121. Scopus. https://doi.org/10.1007/s10796-022-10269-2

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). 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*, *26*(4), 2141–2168. Scopus. https://doi.org/10.1007/s11948-019-00165-5

Murray, D. (1997). *Ethics in Organizations*. Kogan Page Publishers.

Nagbøl, P. R., Müller, O., & Krancher, O. (2021). *Designing a Risk Assessment Tool for Artificial Intelligence Systems: Vol. 12807 LNCS* (p. 339). Scopus. https://doi.org/10.1007/978-3-030-82405-1_32

Narayanan, M., & Schoeberl, C. (2023). *A Matrix for Selecting Responsible AI Frameworks*. Center for Security and Emerging Technology. https://doi.org/10.51593/20220029

Noiseau, P. (2023). Ethics of care and Artificial Intelligence: The need to integrate a feminist normative approach. In B. Prud'homme, C. Régis, G. Farnadi, V. Dreier, S. Rubel, & C. d'Oultremont (Eds.), *Missing links in AI governance* (pp. 344–358). Paris : UNESCO; Montréal : Mila – Québec Institute of Artificial Intelligence.

Obermeyer, Z., Nissan, R., Stern, M., Eaneff, S., Bembeneck, E. J., & Mullainathan, S. (2021). *Algorithmic Bias Playbook*. Chicago Booth. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook

Oliveira, D., & Dalkir, K. (2022). Knowledge Capture for the Design of a Technology Assessment Tool. *14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *2*, 185–192. https://doi.org/10.5220/0011551400003335

Perry, N., Srivastava, M., Kumar, D., & Boneh, D. (2022). *Do Users Write More Insecure Code with AI Assistants?* (arXiv:2211.03622). arXiv. https://doi.org/10.48550/arXiv.2211.03622

Probasco, E. (2022). *A Common Language for Responsible AI*. Center for Security and Emerging Technology.

Qiang, V., Rhim, J., & Moon, A. (2023). No such thing as one-size-fits-all in AI ethics frameworks: A comparative case study. *AI & Society*. https://doi.org/10.1007/s00146-023-01653-w

Reich, B. H., & Benbasat, I. (2000). Factors That Influence the Social Dimension of Alignment between Business and Information Technology Objectives. *MIS Quarterly*, *24*(1), 81–113.

Richards, J., Piorkowski, D., Hind, M., Houde, S., & Mojsilović, A. (2020). *A Methodology for Creating AI FactSheets*. http://arxiv.org/abs/2006.13796

Rolls Royce. (2021). *The Aletheia Framework 2.0*. https://www.rolls-royce.com/~/media/Files/R/RollsRoyce/documents/stand-alone-pages/aletheia-framework-booklet-2021.pdf

Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science and Medicine*, *296*. Scopus. https://doi.org/10.1016/j.socscimed.2022.114782

Skitka, L. J., Mosier, K. L., Burdick, M., & Rosenblatt, B. (2000). Automation bias and errors: Are crews better than individuals? *International Journal of Aviation Psychology*, *10*(1), 85–97. Scopus. https://doi.org/10.1207/S15327108IJAP1001_5

Suresh, H., & Guttag, J. (2021). *A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle*. ACM International Conference Proceeding Series. Scopus. https://doi.org/10.1145/3465416.3483305

Tursunbayeva, A., & Renkema, M. (2022). Artificial intelligence in health-care: Implications for the job design of healthcare professionals. *Asia Pacific Journal of Human Resources*. https://doi.org/10.1111/1744-7941.12325

Université de Montréal. (2018). *Declaration of Montréal for a responsible development of AI*. https://www.montrealdeclaration-responsibleai.com

Vassileva, B. K. (2021). Artificial Intelligence: Concepts and Notions. In B. Vassileva & M. Zwilling (Eds.), *Advances in Human and Social Aspects of Technology* (pp. 1–18). IGI Global. https://doi.org/10.4018/978-1-7998-4285-9.ch001

Verbeek, P.-P. & Tijink, Daniël. (2020). *Guidance ethics approach: An ethical dialogue about technology with perspective on actions*.

von Richthofen, G., Ogolla, S., & Send, H. (2022). Adopting AI in the Context of Knowledge Work: Empirical Insights from German Organizations. *Information*, *13*(4). https://doi.org/10.3390/info13040199

World Economic Forum. (n.d.). *Empowering AI Leadership: An Oversight Toolkit for Boards of Directors*. World Economic Forum. https://express.adobe.com/page/RsXNkZANwMLEf/.