Building Trustworthy and Human-Centered Intelligent Information Systems for a Sustainable Future

Marcos Kalinowski[1], Allysson Allex Araújo[1,2], Simone D. J. Barbosa[1], and Hélio Lopes[1]

[1]Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rio de Janeiro, Rio de Janeiro – Brasil

[2]Universidade Federal do Cariri (UFCA)
Juazeiro do Norte, Ceará – Brasil

{kalinowski, simone, lopes}@inf.puc-rio.br allysson.araujo@ufca.edu.br

Abstract: *Artificial Intelligence (AI) is a multidisciplinary field dedicated to creating systems capable of performing tasks that typically require human intelligence. Intelligent Information Systems (IIS) have been defined as information systems that integrate AI technologies. Given the increasing integration of AI into IIS, a foundational challenge for the next decade lies in ensuring that AI-powered IIS are developed in a way that prioritizes trustworthiness, careful consideration of ethical and human-centered principles, and positive societal impact. This challenge highlights the need to embed Trustworthy AI and Human-Centered AI principles as essential pillars in the design and governance of IIS. Trustworthy AI involves the development of AI systems that offer benefits and convenience while mitigating threats and minimizing risks of harm. Complementarily, Human-Centered AI emphasizes the alignment of AI with core human values, ensuring that these systems are both ethically responsible and socially beneficial. Integrating these paradigms can help to ensure that IIS are designed for a sustainable future..*

Palavras-chave*: Trustworthy AI, Human-Centered AI, Intelligent Information Systems.*

1. What is your idea, vision, or reflection on the challenge for IS in Brazil over the next ten years?

Artificial Intelligence (AI) is a multidisciplinary field dedicated to creating systems capable of performing tasks that typically require human intelligence [Lu 2019]. Intelligent Information Systems (IIS) have been defined as information systems that integrate AI technologies [Ras and Tsay 2010]. Given the increasing integration of AI into IIS, a foundational challenge for the next decade lies in ensuring that AI-powered IIS are developed in a way that prioritizes trustworthiness, careful consideration of ethical and human-centered principles, and positive societal impact.

This challenge highlights the need to embed Trustworthy AI and Human-Centered AI principles as essential pillars in the design and governance of IIS. Trustworthy AI involves the development of AI systems that offer benefits and convenience while mitigating threats and minimizing risks of harm [Liu et al. 2022, Baldassarre et al. 2024b].

Complementarily, Human-Centered AI emphasizes the alignment of AI with core human values, ensuring that these systems are both ethically responsible and socially beneficial [Ozmen Garibay et al. 2023]. Integrating these paradigms can help to ensure that IIS are designed for a sustainable future.

The vision behind this challenge involves promoting the development of IIS that incorporates ethical AI frameworks [Li 2024]. Furthermore, it calls attention to rethinking and expanding IS education to prepare professionals capable of design- ing, deploying, and managing fair, interpretable, and accountable IIS. Beyond educa- tion, this challenge urges the establishment of robust regulatory and technical mecha- nisms to guide AI governance, ensuring transparent decision-making, ongoing audits, and bias mitigation strategies to safeguard social well-being. For instance, while doc- umentation tools such as Model Cards [Mitchell et al. 2019] and Semiotic Engineer- ing's Extended Metacommunication Template [Barbosa et al. 2021] can aid developers in ethical reflection, studies suggest that they may lead to selective reporting of con- cerns, highlighting the need for stronger accountability measures in AI system design [Nunes et al. 2024, Nunes et al. 2022, Barbosa et al. 2024].

This challenge is inherently interdisciplinary, requiring collaborative efforts among academia, industry, government, and civil society to ensure that the transforma- tive potential of AI contributes to the nation's development in ways that are socially just, environmentally sustainable, and economically competitive. By framing this challenge within IS research beyond developing technologically feasible solutions aligned with hu- man values, we can leverage its focus on technological innovation and organizational and societal impact.

2. Why is it critical for the community to address this challenge?

The integration of trustworthy and human-centered AI into IIS is critical for several key reasons. First, the unchecked deployment of AI-powered IIS can exacerbate so- cial inequalities, particularly in a diverse and economically stratified country like Brazil. Without intentional efforts to design and implement IIS that align with human values, marginalized communities may suffer disproportionately from biased algorithms and ar- bitrary decision-making processes [Souza et al. 2023, Nicola´s and Sampaio 2024].

Second, ensuring trust in AI-driven systems is fundamental for widespread adoption and societal acceptance [Baldassarre et al. 2024a]. If IIS are perceived as unfair or unaccountable, public resistance could emerge, undermining their transfor- mative potential in sectors such as healthcare, education, and public administration [Lockey et al. 2021, Robles and Mallinson 2023, Jungherr 2023]. Proactively addressing these concerns is important for ensuring public confidence in AI's role across these do- mains [Mesquita et al. 2024, Glikson and Woolley 2020].

Furthermore, embedding sustainability and ethical considerations into IIS aligns with global commitments, such as the United Nations' Sustainable Development Goals (SDGs) [Borsatto et al. 2024, Vinuesa et al. 2020, Palomares et al. 2021]. By prioritizing explainability, fairness, and ecological responsibility, IIS can become powerful tools for advancing social and environmental progress, ensuring that technology serves as a force for good [Carney 2020, Cath 2018].

3. What are the risks if progress is not made in solving this challenge?

Failure to address the integration of trustworthy and human-centered AI into IIS poses severe ethical, social, and economic risks. For example, without trustworthy and human-centered AI, existing social inequities may be amplified, as biased algorithms disproportionately affect underrepresented groups in areas such as hiring, access to services, and criminal justice [Nicola´s and Sampaio 2024]. This issue could erode trust in technology and exacerbate societal divisions [Liu et al. 2022]. The risks extend beyond inequality, raising concerns about misuse of surveillance technologies, violations of privacy, and manipulation of public opinion, which could further undermine democratic processes and public confidence in AI-driven systems [Ozmen Garibay et al. 2023].

Moreover, Brazil risks falling behind in the global AI innovation race. As international competition intensifies, countries that fail to integrate ethical and human-centered principles into their AI systems will struggle to meet emerging global standards, potentially leading to exclusion from international research and economic collaborations [Vinuesa et al. 2020]. This could result in not only missed economic opportunities but also in increased technological dependence on external actors, threatening Brazil's na- tional sovereignty.

Finally, neglecting sustainability considerations in the design of IIS could hinder Brazil's ability to achieve the SDGs and leave the country ill-prepared to address critical environmental and social challenges [Vinuesa et al. 2020, Palomares et al. 2021]. If IIS are not developed and implemented responsibly, they can contribute to environmental degradation and social inequalities, obstructing progress toward sustainable development. Therefore, integrating trustworthy and human-centered AI frameworks into IIS is relevant to ensure that technological advancements contribute positively to society and the environment [Souza et al. 2023, Mesquita et al. 2024].

4. How does this challenge relate to other problems, areas, knowledge, actions, initiatives, and technologies?

We can begin with data governance and AI auditing mechanisms, which are critical to ensuring the ethical use and protection of sensitive information in our increasingly digital world [Janssen et al. 2020, Fischer and Piskorz-Ryn 2021]. Upholding principles such as transparency, consent, and accountability is important in fostering confidence in IIS and goes beyond technical requirements [Solano et al. 2022].

We also need to rethink IS education, placing a strong emphasis on multidisciplinary subjects like critical thinking, inclusive design, and AI ethics [Memarian and Doleck 2023, Borsatto et al. 2024]. Public policy is an- other important and related topic. Regulatory and propositional frameworks that deal with accountability and prevent the misuse of AI are highly perti- nent [de Almeida et al. 2021, Baldassarre et al. 2024a]. In fact, the implications of integrating trustworthy and human-centered AI extend beyond governance and edu- cation in the context of IIS, intersecting with broader social and environmental goals. By addressing these interconnections and the role of trustworthy and human-centric approaches, we can drive technological progress while ensuring a more equitable and sustainable future for all.

References

Baldassarre, M. T., Gigante, D., Kalinowski, M., and Ragone, A. (2024a). Polaris: A framework to guide the development of trustworthy ai systems. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI, pages 200–210.

Baldassarre, M. T., Gigante, D., Kalinowski, M., Ragone, A., and Tibido`, S. (2024b). Trustworthy ai in practice: an analysis of practitioners' needs and challenges. In Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, pages 293–302.

Barbosa, G. D. J., Nunes, J. L., De Souza, C. S., and Barbosa, S. D. J. (2024). Investigating the Extended Metacommunication Template: How a semiotic tool may encourage reflective ethical practice in the development of machine learning systems. In Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems, IHC '24, pages 1–12, New York, NY, USA. Association for Computing Machinery.

Barbosa, S. D. J., Barbosa, G. D. J., Souza, C. S. d., and Leita˜o, C. F. (2021). A Semiotics-based epistemic tool to reason about ethical issues in digital technology design and development. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, pages 363–374, New York, NY, USA. Association for Computing Machinery.

Borsatto, J. M. L. S., Marcolin, C. B., Abdalla, E. C., and Amaral, F. D. (2024). Aligning community outreach initiatives with sdgs in a higher education institution with artificial intelligence. Cleaner and Responsible Consumption, 12:100160.

Carney, T. (2020). Artificial intelligence in welfare: Striking the vulnerability balance? Monash University Law Review, 46(2):23–51.

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133):20180080.

de Almeida, P. G. R., dos Santos, C. D., and Farias, J. S. (2021). Artificial intelligence regulation: a framework for governance. Ethics and Information Technology, 23(3):505–525.

Fischer, B. and Piskorz-Ryn´, A. (2021). Artificial intelligence in the context of data governance. Interna- tional Review of Law, Computers & Technology, 35(3):419–428.

Glikson, E. and Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2):627–660.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., and Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. Government information quarterly, 37(3):101493.

Jungherr, A. (2023). Artificial intelligence and democracy: A conceptual framework. Social media+ society, 9(3):20563051231186353.

Li, Z. (2024). Ethical frontiers in artificial intelligence: navigating the complexities of bias, privacy, and accountability. International Journal of Engineering and Management Research, 14(3):109–116.

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A., and Tang, J. (2022). Trustworthy ai: A computational perspective. ACM Transactions on Intelligent Systems and Technology, 14(1):1–59.

Lockey, S., Gillespie, N., Holm, D., and Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions.

Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1):1–29.

Memarian, B. and Doleck, T. (2023). Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (ai), and higher education: A systematic review. Computers and Education: Artificial Intelligence, page 100152.

Mesquita, H., Garrote, M. G., and Zanatta, R. A. (2024). Regulating artificial intelligence in brazil: the contributions of critical social theory to rethink principles. Technology and Regulation, 2024:73–83.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. (2019). Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* ’19, pages 220–229, New York, NY, USA. Association for Computing Machinery.

Nicolas, M. A. and Sampaio, R. C. (2024). Balancing efficiency and public interest: The impact of ai automation on social benefit provision in Brazil. Internet Policy Review, 13(3).

Nunes, J. L., Barbosa, G. D. J., De Souza, C. S., and Barbosa, S. D. J. (2024). Using Model Cards for ethical reflection on machine learning models: an interview-based study. Journal on Interactive Systems, 15(1):1–19.

Nunes, J. L., Barbosa, G. D. J., de Souza, C. S., Lopes, H., and Barbosa, S. D. J. (2022). Using model cards for ethical reflection: a qualitative exploration. In Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems, IHC ’22, pages 1–11, New York, NY, USA. Association for Computing Machinery.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., et al. (2023). Six human-centered artificial intelligence grand challenges. International Journal of Human–Computer Interaction, 39(3):391–437.

Palomares, I., Mart´ınez-Ca´mara, E., Montes, R., Garc´ıa-Moral, P., Chiachio, M., Chiachio, J., Alonso, S., Melero, F. J., Molina, D., Ferna´ndez, B., et al. (2021). A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. Applied Intelligence, 51:6497–6527.

Ras, Z. and Tsay, L.-S. (2010). Advances in Intelligent Information Systems.

Robles, P. and Mallinson, D. J. (2023). Artificial intelligence technology, public trust, and effective gover- nance. Review of Policy Research.

Solano, J. L., de Souza, S., Martin, A., and Taylor, L. (2022). Governing data and artificial intelligence for all: models for sustainable and just data governance.

Souza, J., Avelino, R., and da Silveira, S. A. (2023). Artificial intelligence: dependency, coloniality and technological subordination in Brazil. In Elgar Companion to Regulating AI and Big Data in Emerging Economies, pages 228–244. Edward Elgar Publishing.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Fella¨nder, A., Langhans, S. D., Tegmark, M., and Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. Nature communications, 11(1):1–10.