

ARTIFICIAL INTELLIGENCE AND RESPONSIBLE MANAGEMENT – A Framework of Technological, Organizational, and Societal Principles

Farley Simon Nobre, Ph.D.

Associate Professor of Business Sustainability and Artificial Intelligence

PPGADM, Management School, Federal University of Paraná

Av. Prefeito Lothário Meissner 632, 80210-170, Curitiba-PR, Brazil

ORCID: <https://orcid.org/0000-0002-1595-8903>

e-mail: f.nobre@ufpr.br ; fsmnobre@gmail.com

Abstract: There is a growing interest in how artificial intelligence (AI) shapes business and society, with great attention paid to management and organization studies. However, more needs to be said about the scope of AI-responsible management. This article puts forward conceptual relationships between AI, AI designers, AI users/ adopters, and the society targets represented by the United Nations Sustainable Development Goals (SDGs). By setting these relationships, it introduces definitions of AI as intelligent and cognitive machines, designers of AI, the symbolic consciousness of the AI (machine) that performs tasks and participates in the organization, and the work responsibilities that set relationships between AI, AI designers, and the organizations adopting them. It also discusses explainable AI as a necessary feature to improve human trust in AI. Finally, it connects these conceptual AI elements to society and proposes a responsible management framework for AI.

Keywords: Artificial Intelligence, Responsible Management, Explainable AI, Business and Society, Sustainability.

Track: Sustainable and Responsible Business

Word count: 2051 (excluding figures and references)

How to cite this article:

Nobre, F.S. (2024). Artificial Intelligence and Responsible Management – A Framework of Technological, Organizational, and Societal Principles. *British Academy of Management (BAM Conference)*, Track: Sustainable and Responsible Business, pp.1-10. Nottingham Trent University, 02-06 Sept. 2024.

“If one assumes that the cognitive roles in the world of organizations, as fulfilled by agents, have performance and outcomes which can be attributed to either humans or machines, without any distinction, then one is ready to consider machines as members of organizations which act in the name of them similarly to people.”¹ (Nobre, 2005: subsection 1.7.6)

INTRODUCTION

Herbert A. Simon (1916-2001) pioneered research on the impact of computers and artificial intelligence (AI) in management and organizations (Simon, 1960). Following a cycle of AI spring and winter (Mitchell, 2021), today, we entered an era of AI's highest promise, especially with the advent of generative AI (MIT Technology Review Insights, 2023; Noy & Zhang, 2023).

In this context, there is a particularly sharp spike of renewed interest in how AI shapes businesses, managerial and entrepreneurial tasks, and organizations (e.g., Balasubramanian & Ye, 2022; Csaszar & Steinberger, 2022; Glikson & Woolley, 2020; Grimes et al., 2023; Raisch & Fomina, 2023; Raisch & Krakowski, 2021; Townsend & Hunt, 2019; Townsend et al., 2023; von Krogh, Roberson, & Gruber, 2023).

However, there needs to be more management and organization studies that explain the interplay between AI and responsible management. This article addresses this issue by setting relationships between AI (as artifacts/ machines that comprise AI algorithms and other AI types), AI designers (i.e., the firms developing and commercializing AI), AI users/ adopters (i.e., people and the organizations directly and indirectly adopting AI to achieve their motives, purposes, and goals), and the society targets represented by the United Nations Sustainable Development Goals² - SDGs (UN, 2015).

Management tasks and organizations have been challenged and shaped by new organizing processes, advanced information systems, and machines in the form of computers, algorithms, communication networks, and other technologies. When adequately designed with a background in artificial intelligence, cognitive science, and systems theory, these machines can subsume capabilities to carry out intelligent and cognitive organizational tasks (Nobre, 2005; Nobre et al., 2009a). Indeed, these machines have been developed to participate in organizations, performing automation and augmentation processes (Raisch & Fomina, 2023; Raisch & Krakowski, 2021). Hence, they enter the arena of organizations as additional design elements that can support and compete with human beings when performing technical and managerial tasks at similar or superior levels (e.g., Lenharo, 2024b). However, the fast progress in designing and implementing AI needs technological, economic, social, and policy analysis (Nobre, 2005; Vinuesa *et al.*, 2020).

Understanding the new challenges AI machines raise for academics, managers, and entrepreneurs in organizations and society outlines the need for developing novel propositions and organization theories (Grimes et al., 2023; Nobre, 2005; Nobre et al., 2009a, 2009b; von Krogh, Roberson, & Gruber, 2023). I am not pretending to introduce a new theory of AI in management, such as a theory of cognitive machines in organizations (Nobre, 2005, 2012; Nobre et al., 2008; 2009a, 2009b). However, **I aim to outline a responsible management framework for AI**. It shall include and move beyond technical principles of AI explainability and security (Arrieta et al., 2020). My framework shall present the flow of relationships between AI, AI designers, AI

¹ This analysis is built on the classical imitation game (Turing, 1950).

² The SDGs provide a coherent and holistic framework for creating a better future based on collaborative efforts by all parties in society, including businesses. They are organized into the Universal 2030 Agenda, which comprises 17 SDGs and 169 related targets to address the world's most urgent economic, social, environmental, and governance challenges in 2015-2030 (UN, 2015).

adopters, and society. **Further research on this article will be developed through an integrative review of AI and management studies.**

PRELIMINARY DEFINITIONS

Cognitive Machines

Definition 1 (Nobre et al., 2009b: 72): “Cognitive machines are agents whose functioning processes are mainly inspired by human cognition. Therefore, they have great possibilities to present intelligent behavior.”

Artificial Intelligence (AI) Machines

Definition 2 (Glikson & Wooley, 2020: 627): “Artificial intelligence (AI) represents a highly capable and complex technology that aims to simulate human intelligence.”

Figure 1 illustrates the design strategies for machine cognition and intelligence.

Insert Figure 1 about here.

There are other definitions for AI with more variables and complexities (Csaszar & Steinberger, 2022; Russell & Norvig, 2020). Despite the theoretical distinction between cognitive and AI machines (Nobre, 2005: subsection 3.2; Nobre et al., 2009b: 70), I will use them as synonymous in this article, which aims to outline a responsible management framework for AI.

AI and Cognitive Machines Designers

Definition 3 (Nobre et al., 2009a: 375 – with proofreading): “The designer of AI or cognitive machine is the person (or third-party organization) responsible for the machine’s cognitive abilities and intelligent behavior. AI designers can have an independent legal identity, enabling them to make contracts and seek court enforcement if necessary.”

Definition 4 (Nobre et al., 2009a: 375 – with proofreading): “AI designers/ engineers can use AI technologies to design and engineer AI or cognitive machines automatically³. Such technologies are classified as artificial intelligence designers and cannot ask for or answer a formal contract. Therefore, their first designer (a human or third-party organization) is the agent who can make it.”

³ Genetic programming for instance is a paradigm which has provided a profound impact on the design of software programs capable of generating tangible replicas and with enough ability to perform at least similar functions (Koza 1992).

AI and Cognitive Machines' Consciousness in the Organization

Consciousness regards our mental states of being aware of ourselves. It develops from our sensory experience and aspects of proprioception and introspection (Barsalou, 1999). It assumes the awareness of our mental processes, thoughts, feelings, and perceptions. Consciousness states can vary from deep sleep to alert wakefulness (Bernstein et al., 1997). Indeed, there are competing theories for how the brain produces an individual's subjective experience that enters consciousness. Some favor artificial intelligence, and others reject it (Lenharo, 2024a).

This article shall not discuss whether AI or cognitive machines can reach consciousness and develop awareness about their organizational tasks and participation. However, I defined what (AI) machines' consciousness means for the organizations.

Definition 5 (Nobre et al., 2009a: 376 – with proofreading): “AI machine's consciousness represents the awareness of its designer about the cognitive abilities and intelligent behavior that the AI machine can perform during task execution.”

RELATIONSHIPS BETWEEN AI, AI DESIGNERS, AND AI ADOPTERS

Regarding responsible management of the relationships between AI, AI designers, and AI users/adopters (the organizations), I suggest regulatory definitions that promote transparency of their obligations.

Definition 6 (Nobre et al., 2009a: 376 – with proofreading): “The working relationship between the AI designer/ producer and the organizations/ AI adopters shall be defined by a legal contract that makes explicit the AI cognitive abilities, the AI behavioral expectations, the AI tasks and roles in the organization, and an attestation/ signature of both parts.”

Definition 7 (Nobre et al., 2009a: 376 – with proofreading): “The organization/ AI adopter is responsible for assigning roles to the AI machine, and the AI machine is responsible for the roles it assumes in the organization. However, the AI designer and the organization/ adopter are the main agents responsible for the AI behavior and performance. Suppose the AI exhibits deviant behavior during task execution or underperforms below specified criteria. In that case, the contract between the organization/ adopter and the AI designer is the object of analysis and judgment.”

AI TRANSPARENCY AND EXPLAINABILITY

Most AI algorithms are conceived as opaque or sub-symbolic black-box systems since they rely on deep learning neural networks, big data, and probabilistic and statistical computation (Castelvecchi, 2016; Csaszar & Steinberger, 2022; Townsend et al., 2023). Despite presenting better outcomes in many applications than other AI symbolic systems, these types of AI fail to offer explainability and transparency, i.e. a trade-off between performance and interpretability (Arrieta et al., 2020; Garibaldi, 2019; Gunning et al., 2019; Hagendorff & Wezel, 2020) and can lead to distrust (Glikson & Woolley, 2020), causing barriers to achieve what researchers propose as the cycle of human-AI augmentation (Raisch & Fomina, 2023; Raisch & Krakowski, 2021; Wang et al., 2022). Accounting for AI processes (e.g., perception, categorization, decision-making, and learning) is necessary to make AI transparent to designers and adopters. In the search

for AI transparency, some authors have proposed interpretable and explainable AI algorithms, theories, and applications in deep learning (Samek et al., 2019). Others have developed new directions in AI (Zadeh, 2001) built on computational and cognitive models (Garibaldi, 2019; Gu, 2023; Nobre, 1995, 2005; Nobre *et al.*, 2009a, 2009b; Wang, 1994; Zadeh, 1994) that overcome the lack of transparency and interpretability in most probabilistic and statistical AI systems. These alternative AI models are complementary rather than substitutive of probabilistic reasoning (Zadeh, 2006). They rely on linguistic categories and natural language to exploit the remarkable ability of the human brain to solve ill-defined and abstract problems and reach approximate solutions without quantifiable measurements (Garibaldi, 2019; Nobre, 2005; Zadeh, 1999, 2015).

Definition 8: AI designers shall develop AI to meet interpretable and explainable criteria. In addition to the AI designers, the organizations/ AI adopters shall be able to interpret and explain the AI algorithms' behavioral and cognitive processes.

RESPONSIBLE RELATIONSHIPS BETWEEN AI AND SOCIETY

Regarding the interplay between AI and society, some authors have mapped and discussed the impact of AI on the 17 SDGs and their 169 targets (Vinuesa *et al.*, 2020). Findings suggested that AI can favor the progress of 134 targets. However, they may also hinder 59 targets (*ibid*: 1). The authors argued the need for regulatory insight and oversight for AI to promote social, environmental, and economic sustainable development. Having said that, I propose:

Definition 9: AI designers and AI adopters shall develop and apply AI to meet social, environmental, and economic goals that promote sustainable development.

Figure 2 introduces a responsible management framework that comprises the flow of relationships between AI, AI designers, AI adopters, and society.

Insert Figure 2 about here.

CONCLUSION

Public and private institutions, especially all forms of profit and non-profit business organizations, that develop, make available, commercialize, or adopt AI should serve society and safeguard the planet by attending responsible management principles that regulate relationships, transparency, and sustainable development. Like the purpose of the principles for responsible management education (PRME, 2024), *responsible AI* seeks to guide people who will help their organizations create inclusive prosperity while promoting transparency, freedom, justice, and peace within regenerative and resilient natural ecosystems. AI organizations must follow truly sustainable businesses (Dyllick & Muff, 2016; Nobre, 2024) since they represent powerful actors that can favor sustainable development and well-being. Indeed, without a responsible management framework, AI organizations are subject to blind and error-prone strategic decisions oriented by short-termism and the hope that they are on the right path, but without the necessary knowledge to correct their course of action in the long run to foster social, environmental, and economic sustainable development.

REFERENCES

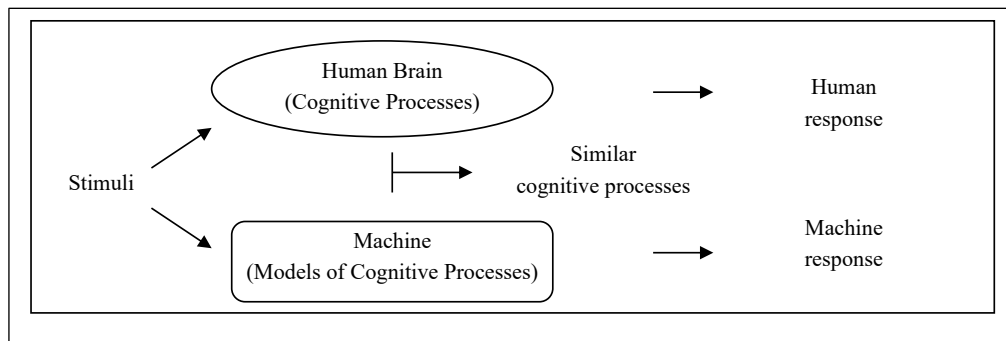
- Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
- Balasubramanian, N., & Ye, Y. (2022). Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. *Academy of Management Review*, 47(3), 448-465.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Science*, 22, 577-660.
- Bernstein, D., Penner, L.A., Clarke-Stewart, & Roy, E. (2007). *Psychology*. Houghton Mifflin Company: Boston.
- Castelvecchi, D. (2016) Can we open the black box of AI? *Nature*, 538(7623), 20–23.
- Csaszar, F.A., & Steinberger, T. (2022). Organizations as Artificial Intelligences: The Use of Artificial Intelligence Analogies in Organization Theory. *Academy of Management Annals*, 16(1), 1-37.
- Dyllick, T., & Muff, K. (2016). Clarifying the meaning of sustainable business: Introducing a typology from business-as-usual to true business sustainability. *Organization & Environment*, 29(2), 156-174.
- Garibaldi, J.M. (2019). The Need for Fuzzy AI. *IEEE/CAA Journal of Automatica Sinica*, 6(3), 610-622.
- Glikson, E., & Woolley, A.W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, 14(2), 627-660.
- Grimes, M., von Krogh, G., Feuerriegel, S., Rink, F., & Gruber, M. (2023). From the Editors. From Scarcity to Abundance: Scholars and Scholarship in an Age of Generative Artificial Intelligence. *Academy of Management Journal*, 66(6), 1617–1624.
- Gu, X., Han, J., Shen, Q., & Angelov, P.P. (2023). Autonomous learning for fuzzy systems: a review. *Artificial Intelligence Review*, 56, 7549-7595.
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. & Yang, G-Z. (2019). XAI- Explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120.
- Hagendorff, T., & Wezel, K. (2020) 15 challenges for AI: or what AI (currently) can't do. *AI & Society*, 35, 355–365. <https://doi.org/10.1007/s00146-019-00886-y>
- Lenharo, M. (2024a). Consciousness: The Future of an Embattled Field. *Nature*, 628, 438-440.
- Lenharo, M. (2024b). Google AI has better bedside manner than human doctors — and makes better diagnoses. *Nature*. 12 January 2024. <https://doi.org/10.1038/d41586-024-00099-4>
- MIT Technology Review Insights (2023). *The great acceleration: CIO perspectives on generative AI*. Retrieved Jan. 20, 2024. https://www.databricks.com/sites/default/files/2023-07/ebook_mit-cio-generative-ai-report.pdf
- Mitchell, M. (2021). Why AI is harder than we think. Ithaca, NY: arXIV, Cornell University. <https://doi.org/10.48550/arXiv.2104.12871>
- Nobre, F.S. (1995). Genetic-Neuro-Fuzzy Systems: A Promising Fusion. *IEEE International Conference on Fuzzy Systems - FUZZ-IEEE/IFES*, Yokohama, Japan, 20-24 March 1995. doi: 10.1109/FUZZY.1995.409690
- Nobre, F.S. (2005). *On Cognitive Machines in Organizations*. Ph.D. Thesis, University of Birmingham, Birmingham Main Library, Control Number M0266887BU, Birmingham-UK. <https://etheses.bham.ac.uk/id/eprint/12651/>
- Nobre, F.S. (2012). Governing Industrial Organizations through Cognitive Machines. *AI & Society*, 27(4):501-507. <https://doi.org/10.1007/s00146-011-0364-0>

- Nobre, F.S. (2024). Unleashing Virtuous Cycles of Sustainable Development Goals and Well-Being. *Business and Society Review*. Special Issue: Moving the Logic of Sustainability Towards Flourishing for All. Accepted for publication (25 October 2023). DOI: 10.1111/BASR.12339
- Nobre, F.S., Tobias, A.M. & Walker, D.S. (2008). The Pursuit of Cognition in Manufacturing Organizations. *Journal of Manufacturing Systems*, 27(4), 145-157. <https://doi.org/10.1016/j.jmsy.2009.01.001>
- Nobre, F.S., Tobias, A.M. & Walker, D. (2009a). The Impact of Cognitive Machines in Complex Decisions and Organizational Change. *AI & Society*, 24(4): 365-381. <https://doi.org/10.1007/s00146-009-0207-4>
- Nobre, F.S., Tobias, A.M. & Walker, D. (2009b). *Organizational and Technological Implications of Cognitive Machines: Designing Future Information Management Systems*. New York: IGI Global. <http://doi.org/10.4018/978-1-60566-302-9>
- Noy, S. & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.
- PRME (2024). *The Principles for Responsible Management Education*. United Nations. Retrieved Jan. 21, 2024. <https://www.unprme.org/prme-principles-refresh/>
- Raisch, S., & Fomina, K. (2023). Combining Human and Artificial Intelligence: Hybrid Problem-Solving in Organizations. *Academy of Management Review*, 0(0), 1-24.
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review*, 46(1), 192-210.
- Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Hoboken, NJ: Pearson.
- Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., & Muller, K.-R. (Eds.). (2019). *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. *Lecture Notes in Computer Science*. Springer Nature, Switzerland.
- Simon, H.A. (1960). *The new science of management decision*. New York: Harper & Row.
- Townsend, D. M., & Hunt, R.A. (2019). Entrepreneurial action, creativity, & judgment in the age of artificial intelligence. *Journal of Business Venturing Insights*, 11, e00126.
- Townsend, D.M., Hunt, R.A., Rady, J., Manocha, P., & Jin, J.H. (2023). Are the Futures Computable? Knightian Uncertainty and Artificial Intelligence. *Academy of Management Review*, 0(0), 1-26.
- Turing A.M. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433–460.
- UN (2015). *Transforming our world: The 2030 agenda for sustainable development*. A/RES/70/1. United Nations: Geneva. Retrieved March 19, 2020. <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- Vinuesa, R., Azizpour, H., Leite, I. *et al.* (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(233), 1-10. <https://doi.org/10.1038/s41467-019-14108-y>
- von Krogh, G., Roberson, Q., & Gruber, M. (2023). From the Editors. Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence. *Academy of Management Journal*, 66(2), 367-373.
- Wang, B., Zheng, P., Yin, Y., Shih, A., & Wang, L. (2022). Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective. *Journal of Manufacturing Systems*, 63, 471–490.

- Wang, L. (1994) *Adaptive fuzzy systems and control: design and stability analysis*. Englewood Cliffs: PTR Prentice-Hall.
- Zadeh, L.A. (1994). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*, 37(3):77–84.
- Zadeh, L.A. (1999). From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions. *IEEE Transactions on Circuits and Systems*, 45(1), 105-119.
- Zadeh, L.A. (2001). A New Direction in AI: Toward a Computational. Theory of Perceptions. *AI Magazine, Spring*, 73-84.
- Zadeh, L.A. (2015). Fuzzy logic—a personal perspective. *Fuzzy Sets and Systems*, 281, 4–20.

Figure 1. Design strategies focusing on machine cognition and machine intelligence.
Adapted from Nobre (2005).

(a) Machine cognition strategy



(b) Machine intelligence strategy

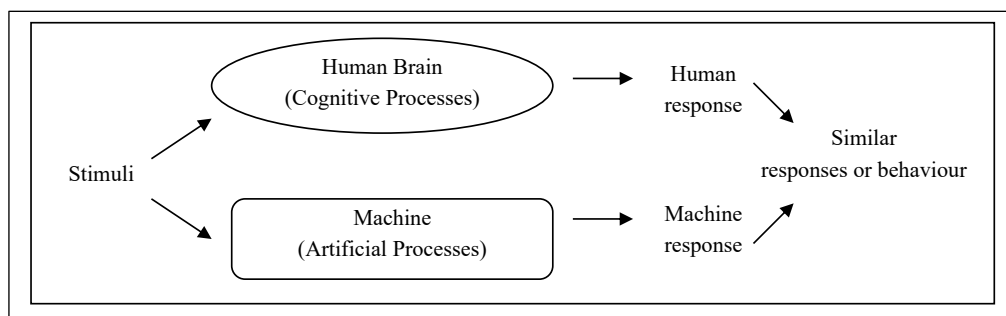


Figure 2. Responsible Management Framework for AI - relationships between AI, AI designers, AI adopters, and society.

