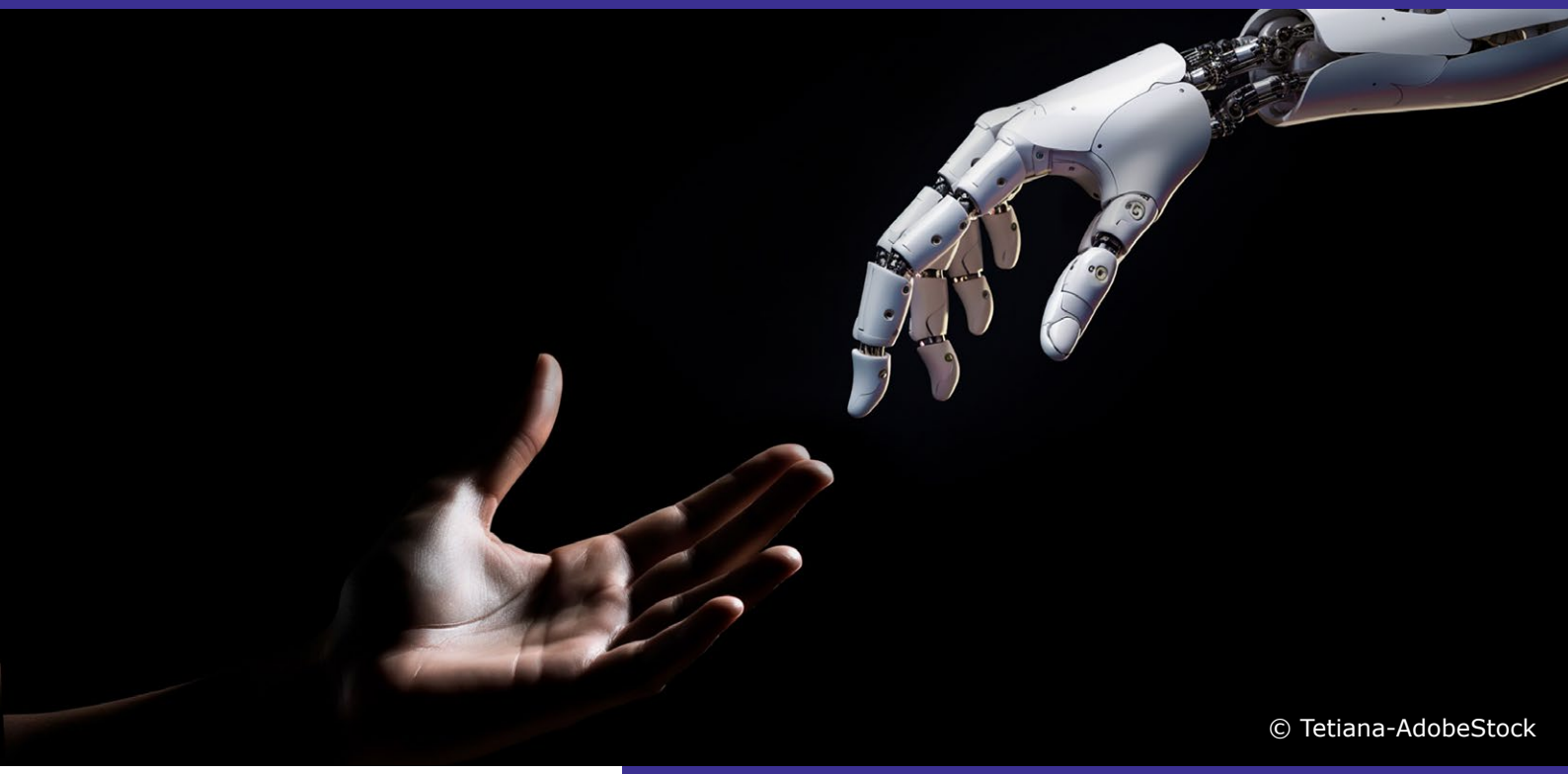
ESCP
BUSINESS SCHOOL

IT ALL STARTS HERE

BERLIN | LONDON | MADRID | PARIS | TURIN | WARSAW



© Tetiana-AdobeStock

LIGHTS - Technology

AI in business: The ART of navigating the tensions between business and social objectives

ESCP Impact Paper No.2024-49-EN

Erik HERMANN, Laetitia MIMOUN & Lorena BLASCO-ARCAS
ESCP Business School

ESCP RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ESCP Impact Paper No.2024-49-EN

AI in business: The ART of navigating the tensions between business and social objectives

Erik Hermann*
Laetitia Mimoun*
Lorena Blasco-Arcas*

ESCP Business School

Abstract

While the advances and benefits of artificial intelligence (AI) for business are indisputable, AI design, development, and deployment do not come without ethical concerns and controversies. Thus, companies have to balance business (e.g., efficiency, personalization, customer satisfaction, profits) on the one hand and corporate digital responsibility and socio-ethical objectives (e.g., privacy, autonomy, sustainability, socially good outcomes) on the other hand. Based on paradox theory, this paper lays out the ART (i.e., Actors, Resources, Time) framework of tension management to provide guidance on how firms and managers can increase the likelihood of eventually turning trade-offs between competing objectives into win-win situations and shared value creation.

Keywords: artificial intelligence, AI ethics, AI for social good, paradox theory, tension management

*Professor, ESCP Business School

AI in business: The ART of navigating the tensions between business and social objectives

Introduction

The design, development, and deployment of artificial intelligence (AI) systems and applications constitute one key organizational challenge across diverse business and industry contexts. Quite recently, this matter has become a key aspect of policy-making, with different government organizations developing guidelines and regulations for better and more sustainable AI implementation and strategic competitiveness in key economy industries. For example, the EU has recently approved the AI Act, intending to address the risks of AI implementation and to position Europe as a leading player globally in AI development. Considering the increasing and fast technical advancement and the potential for value creation, companies and organizations currently are in the race to deploy and implement these systems in their business to benefit from their potential. However, despite all promises, AI implementation has shown to be challenging due to the intra-organizational resistance to its adoption and its potentially harmful intra- and extra-organizational outcomes, raising ethical concerns and calls to responsible management of these technologies for both positive business and societal impact.

One of the critical aspects for businesses to manage AI responsibly is reconciling business objectives with socially good outcomes. Private, public, and research institutions have issued many guidelines and frameworks, including the AI-for-social-good perspective (Cowls et al. 2021; Floridi et al. 2018), to advocate and stipulate ethically and socially responsible AI development and deployment. However, managing these potential tensions and trade-offs between business objectives and social impacts of AI in real business contexts requires moving away from an entirely principled approach to applied ethics and more practical guidance, from principles to practice.

Against this backdrop, our objective is to explore central tensions between business and social objectives by identifying the socio-ethical issues of AI at individual, organizational, and societal levels and laying out a framework and recommendations for companies to manage such tensions.

Socio-ethical issues of AI in business

Despite the various benefits of harnessing AI in business (to pursue business objectives), including automating business tasks and functions, increased efficiency and cost reduction, personalization, customization, and targeting (Brynjolfsson et al. 2023; Huang and Rust 2021; Kunz and Wirtz 2024; Puntoni and Wertenbroch 2024), deploying AI in business does not come without socio-ethical issues and challenges that can oppose social objectives. Accordingly, the rapidly increasing utilization of AI and its encompassing impact at the individual, economic, and societal levels have sparked the discussion of guiding ethical principles and values (e.g., Floridi et al. 2018; Jobin et al. 2019). The effect of AI on society as a whole is central to the AI-for-social-good (Cowls et al. 2021; Floridi et al. 2018), and it thus propagates that AI should be developed and deployed to promote social good (*beneficence*) while preventing any harm (*non-maleficence*; Cowls et al. 2021; Floridi et al. 2018). To achieve such a dual advantage for society, this AI ethics framework entails the ethical principles of *justice*, *autonomy,* and *explicability* in addition to *beneficence* and *non-maleficence* (Floridi et al. 2018). Building on this, we delineate focal socio-ethical issues

across different stakeholder levels based on the AI-for-social-good perspectives and its constituent principles (see Figure 1).

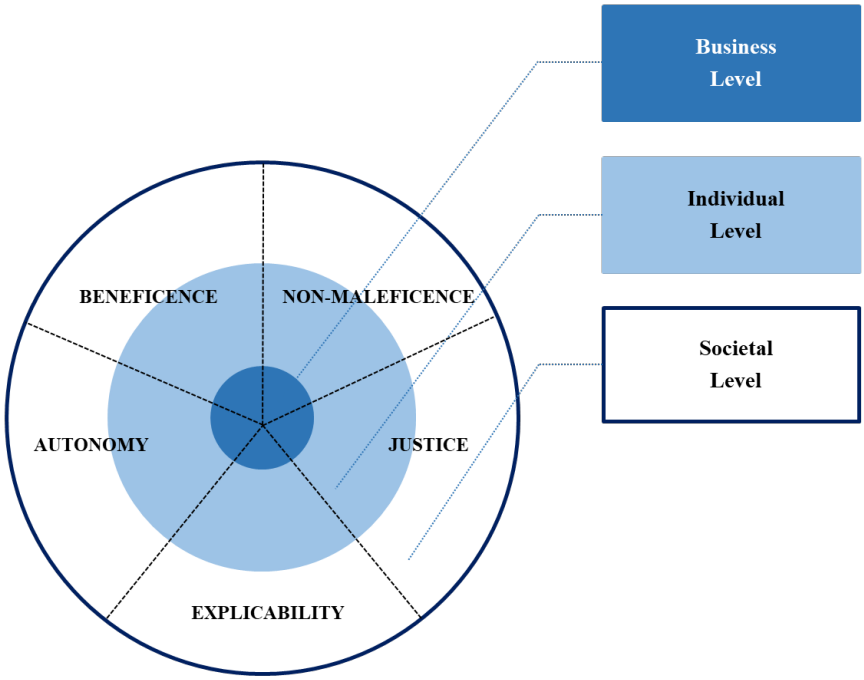


Figure 1. Multi-stakeholder perspective on AI ethics (adapted from Hermann 2022)

First, the environmental impact of AI systems and applications can have adverse societal and (indirect) individual consequences (van Wynsberghe 2021), which implies interferences with the *beneficence* principle. On an individual level, companion AIs have been shown to neither recognize nor appropriately respond to users' mental health issues (De Freitas et al. 2023), undermining *beneficence* as well.

Second, the utilization of AI in business can lead to job (human) replacement (Huang and Rust 2021), which could harm the human workforce and thus infringes the *non-maleficence* principle at the business, individual, and societal levels. Moreover, a trade-off between personalization and privacy can emerge (Rust 2020). That is, the required data inputs of AI applications can interfere with privacy and data protection (Kopalle et al. 2022), which pertain to the individual level and are central to the *non-maleficence* principle (Floridi et al. 2018; Morley et al. 2020). In addition to privacy concerns, data inputs are also decisive for AI systems' inferences and predictions. Broadly speaking, predictions' unbiasedness, validity, and accuracy rest upon input data quality, integrity, and representativeness (Barredo Arrieta et al. 2020; Morley et al. 2020;). Biased predictions can impair *non-maleficence* and *justice (*often used interchangeably with *fairness)* principles at both the individual and societal levels. For instance, discriminatory targeting, segmentation, and prioritization based on demographic, psychological, and economic factors can reinforce existing inequalities, marginalization, biases, stereotypes, and vulnerabilities (Celiktutan et al. 2024; Hermann et al. 2023). Furthermore, recent advances in GenAI are related to deception risks, deep fakes, and manipulative personalization and targeting (Campbell et al. 2022; Schmitt and Flechais 2023), undermining the *non-maleficence principle* at the individual and societal levels.

Third, AI systems and applications shape human agency, autonomy, and self-determination at the business and individual levels (André et al. 2018). On the one hand and at the business level, automation and human enhancement through AI affect workers' autonomy (Puntoni and Wertenbroch 2024). On the other hand and at the individual level, consumers regularly delegate their decisions to AI systems like recommender systems that make suggestions of

what they might like and thus pre-filter information (Kopalle et al. 2022). Overdependence on algorithmic recommendations can, in turn, harm consumer well-being (Banker and Khetani 2019).

Fourth, the opacity and black-box nature of AI systems and applications (Barredo Arrieta et al. 2020) can undermine the *explicability* principle and its epistemological dimension *intelligibility* (Floridi et al. 2018). Business and individual users often lack a basic understanding of AI models' functioning (i.e., *intelligibility*), not to mention the underlying data processing algorithms and models. However, *intelligibility* can lay the foundation for users' ability to judge whether AI systems and their predictions are beneficial (i.e., *beneficence*), harmful (i.e., *non-maleficence*), biased (i.e., *justice*), and whether they should delegate their work and decisions to AI systems (i.e., *autonomy*). Likewise, users and stakeholders, in general, must be aware that they are interacting with artificial agents to draw conclusions about potentially unethical or adverse outcomes (Mozafari et al. 2022).

To address how companies and organizations can manage the tensions between business and social objectives when designing, developing, and deploying AI, we shed light on tension management approaches and Paradox Theory.

Tension Management and Paradox Theory

Companies handle and cope with tensions differently. First, the tensions can be considered as trade-offs or through a win-lose lens, that is, the exchange of a loss for a benefit (Hahn et al. 2010; Van der Byl and Slawinski 2015). Second, they can be treated as paradoxes that are contradictory yet interrelated demands that companies need to accept and continuously manage, and for which there is no single equilibrium that maximizes both social and business goals (Smith and Lewis 2011; Van der Byl and Slawinski 2015). Finally, the win-win paradigm (i.e., the so-called business case) implies the alignment of goals and that the improvement in one goal dimension should improve (or at least not diminish) the other goal dimension, aka shared value creation (Siltaloppi et al. 2021; Van der Byl and Slawinski 2015).

Within the scope of our paper, we adopt a paradox perspective that "explores how organizations can attend to competing demands simultaneously" and argues that the long-term success of an organization "requires continuous efforts to meet multiple, divergent demands" (Smith and Lewis 2011, p. 381). In paradox theory, contradictory elements are not seen as static but interact dynamically, influencing each other over time. According to paradox theory, competing demands reside in dynamic, persistent tensions (Waldman et al. 2019). Rather than aiming to align socio-ethical objectives with business objectives to eliminate tensions, a paradox perspective encourages strategies that accept tensions and attend to different objectives simultaneously, even if they are conflicting. Thus, actors must "embrace the tensions between goals, despite the discomfort associated with juxtaposing opposites, rather than resisting or avoiding tensions" (Van der Byl and Slawinski 2015, p. 59).

Tension management, according to paradox theory, does not rely on choosing between opposing alternatives but rather on managing paradoxes through strategic agility, that is, to rapidly respond to changing environments, to constantly sense and seize opportunities, and to change strategic direction and resource allocation if necessary (Smith and Lewis 2011; Van der Byl and Slawinski 2015). In considering how to do so in the context of AI implementation, we propose that companies manage tensions between business and social objectives by looking at the interactions between actors, resources, and time. Hence, we suggest the ART framework described in the next section.

The ART of tension management

The first and often pivotal step is to recognize a) tensions between objectives and b) the paradoxical nature of them. Our analysis of socio-ethical issues of AI identified several tensions between business and social objectives (i.e., follow ethical principles to achieve socially good outcomes) managers need to address when designing, developing, and implementing AI: (1) Tensions between AI explainability and the protection of proprietary data; (2) Tensions between automation efficiency and human agency and autonomy; (3) Tensions between personalization and targeting and data bias; (4) Tensions between the balance between business growth and AI sustainable implementation and impact.

After becoming aware of paradoxical tensions, we recommend companies and organizations achieve effective and efficient tension management to

a) integrate and account for the perspectives of the different **actors** (stakeholders) involved and affected by AI design and deployment
b) allocate and utilize **resources** in a targeted and flexible manner
c) show long-term commitment, be patient, and seek dynamic equilibria where priorities can shift over **time**

These three recommendations are the building blocks of our ART framework: **A**ctors, **R**esources, and **T**ime (see Figure 2). In the following, we focus on one exemplary tension related to each of the five ethical principles of the AI-for-social good perspective.

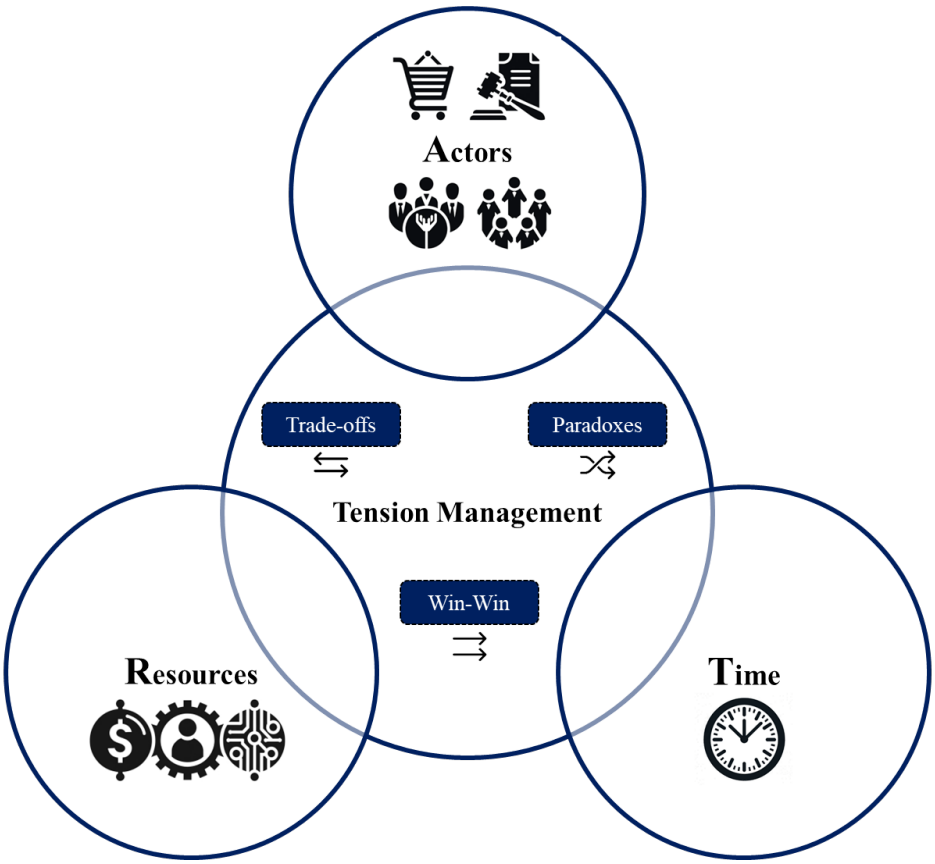


Figure 2. ART framework of tension management

Intelligibility: Understanding AI

Companies and organizations can face tensions between AI systems and applications to be understandable to users and external stakeholders while protecting sensitive and proprietary information and intellectual property. A tiered approach to explaining AI functionalities should be contemplated to manage these tensions. That is, explanations to users should be more simple and basic to avoid irritation, while regulators are (have to be) provided in-depth explanations and information. Therefore, companies and organizations have to invest resources (including human resources such as AI developers, compliance departments, customer experience managers, etc.) and time to create (educational) materials containing explanatory information or user-friendly AI interfaces providing such information. To keep pace with the rapid technological advances of AI and changing regulatory environments, explanations must be constantly scrutinized to ensure understanding across different stakeholders.

Autonomy: Human agency and oversight

In order to manage the tension between efficiency and automation and human autonomy and agency, AI design and development have to prefer human augmentation /enhancement over human replacement. That is, humans have to be kept in the loop - both internally (i.e., workers, employees) and externally (i.e., users) - or in other words, the final decision authority should be left to humans. Developing respective AI systems and applications should involve not only AI developers and managers but also ethicists, legal experts, and end-users (i.e., co-design) to ensure a sufficient level of human agency and self-determination, as well as the human-centric design of AI applications (Blasco-Arcas and Lee 2021). Furthermore, translating ethical principles and guidelines into practices requires nurturing internal expertise and sense-making. The more sophisticated and human-like AI systems and applications become, the more complicated this endeavor might become, which again requires constant assessments over time and respective resource investments to ensure human autonomy.

Justice: Fairness and equity

Companies and organizations should institutionalize auditing and impact assessment mechanisms and bias mitigation and prevention strategies along the entire AI lifecycle to avoid biases, discrimination, and unfair treatment and strengthen data representativeness simultaneously. Companies must allocate technical, human, and financial resources to implement such mechanisms and curate high-quality data sets. Therefore, collaboration with potentially affected (marginalized and discriminated) user groups is necessary to understand better and incorporate their needs and perspectives. Further, companies need to understand at which stages of the consumer journey users may experience vulnerabilities due to the interaction with AI as, in many cases, potential discrimination and bias are contextual and not necessarily structural (Sudbury-Riley et al. 2024). As the diversity of AI applications permanently increases, so do sources of inaccuracies and biases. Thus, continuous feedback and adaptations on AI design and deployment to address emerging justice and fairness concerns are unavoidable.

Non-Maleficence: Human augmentation instead of replacement

As mentioned above, human augmentation should be prioritized to foster autonomy and limit the harm of human and job replacement. Put differently, AI systems and applications enhance employee and worker capabilities and productivity and create new job opportunities instead of replacing jobs. To equip workers and employees with the advanced (technology) skill set and enable them to transition to new roles, companies and organizations have to invest in and provide internal and external (together with educational institutions) training and upskilling programs. Again, AI systems and applications change rapidly and diversify, which requires constant monitoring and adaptation of the required skills sets and respective training programs.

Beneficence: Sustainability

One of the most pivotal and pressing challenges is establishing both AI for sustainability and sustainable AI. In order to leverage AI for business (growth) while managing environmental impacts and promoting sustainability, AI systems and applications should limit resource use (particularly energy and water) and contribute to sustainable development, for instance, by fostering more sustainable consumption behavior. This requires long-term and constant engagement of multiple stakeholders (e.g., environmental organizations and public policy makers) and substantial resource investments (e.g., research and development).

Conclusion

Managing tensions in simultaneously fulfilling business and social impact objectives has become critical for companies designing, developing, and deploying AI systems in current business contexts. Our paper identifies socio-ethical issues that emerge as central organizational tensions and provides a framework for managers to address them. In moving from guidelines to practice, our ART framework provides managers with a tool to redesign AI-related business processes from the perspective of the involved actors, required resources, and time, allowing a more precise understanding to manage those tensions successfully.

References

André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., Huber, J., van Boven, L., Weber, B., & Yang, H. (2018). Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions*, 5(1–2), 28–37.

Banker, S., & Khetani, S. (2019). Algorithm overdependence: How the use of algorithmic recommendation systems can increase risks to consumer well-being. *Journal of Public Policy & Marketing*, 38(4), 500–515.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Benneto, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.

Blasco-Arcas, L., & Lee, H.-H. M. (2021). AI, friend or foe? A proposal for a human-centered approach to AI business implications. *ESCP Impact Paper Series*, No. 2021-07.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *NBER Working Paper Series*, Working Paper 31161.

Campbell, C. Plangger, K., Sands, S., Kietzmann, J. H., & Bates, K. (2022), How deepfakes and artificial intelligence could reshape the advertising industry. *Journal of Advertising Research*, 62(3), 241-251.

Celiktutan, B., Cadario, R., & Morewedge, C. K. (2024). People see more of their biases in algorithms. *Proceedings of the National Academy of Sciences*, 121(16) Article e2317602121.

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, 3(2), 111–115.

De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., & Puntoni, S. (2023). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology*. Advance online publication, https://doi.org/10.1002/jcpy.1393.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.

Hahn, T., Figge, F., Pinkse, J., & Preuss, L. (2010). Trade-offs in corporate sustainability: You can't have your cake and eat it. *Business Strategy and the Environment*, 19(4), 217–229.

Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good - An ethical perspective. *Journal of Business Ethics*, 179(1), 43–61.

Hermann, E., Williams, G. Y., & Puntoni, S. (2023). Deploying artificial intelligence in services to AID vulnerable consumers. *Journal of the Academy of Marketing Science*. Advance online publication. https://doi.org/10.1007/s11747-023-00986-8.

Huang, M.-H., & Rust, R. T. (2021b). A Strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, 39(2), 522–540.

Kunz, W. H., & Wirtz, J. (2023). Corporate digital responsibility (CDR) in the age of AI: implications for interactive marketing. *Journal of Research in Interactive Marketin*g, 18(1), 31-37.

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.

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot – Repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management*, 33(2), 221–245.

Puntoni, S., & Wertenbroch (2024). Being human in the age of AI. *Journal of the Association of Consumer Research. Advance online publication*, https://doi.org/10.1086/730788.

Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing*, 37(1), 15–26.

Schmitt, M., & Flechais I. (2023). Digital deception: Generative artificial intelligence in social engineering and phishing. *SSRN*, http://dx.doi.org/10.2139/ssrn.4602790.

Siltaloppi, J., Rajala, R., & Hietala, H. (2021). Integrating CSR with business strategy: A tension management perspective. *Journal of Business Ethics*, 174(3), 507–527.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review,* 36(2), 381–403.

Sudbury-Riley, L., Hunter-Jones, P., Al-Abdin, A., & Haenlein, M. (2024). When the road is rocky: Investigating the role of vulnerability in consumer journeys. *Journal of the Academy of Marketing Science*. Advance online publication, https://doi.org/10.1007/s11747-024-01011-2.

Van der Byl, C. A., & Slawinski, N. (2015). Embracing tensions in corporate sustainability: A review of research from win-wins and trade-offs to paradoxes and beyond. *Organization & Environment*, *28*(1), 54-79.

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics,* 1(3), 213–218.

Waldman, D. A., Putnam, L. L., Miron-Spektor, E., & Siegel, D. (2019). The role of paradox theory in decision making and management research. *Organizational Behavior and Human Decision Processes*, 155(November), 1-6.