

AI for (un)Sustainability

by Maria Raffa

Abstract

The aim of this paper is to critically examine the presumed benefits of AI in the pursuit of sustainable development. In discussions on today's increasingly urgent global challenges, new technologies – particularly AI – are often hailed as essential solutions. However, there is a significant risk of uncritically embracing AI as a panacea for sustainability without fully considering its disruptive impact across social, ethical, political, and environmental dimensions (Floridi, 2022). In particular, concerns arise regarding the substantial energy consumption of machine learning models and the disposal costs associated with AI-driven devices, such as robots. To address these issues, this paper is structured as follows: it begins with an overview of the current applications of machine learning in sustainability efforts, followed by an analysis of concrete examples where these models contribute to the Sustainable Development Goals (SDGs) established by the United Nations General Assembly in 2015, including cases involving AI-powered robots. The discussion then shifts to the critical challenges posed by these technologies, highlighting how AI itself may threaten sustainability. Finally, a potential model for sustainable AI is proposed, leading to the concluding remarks.

Keywords: Sustainability, AI, Robotics, Techno-solutionism, Environmental Impact

Introduction

In the debate about the global challenges facing our planet in these critical years, there are repeated calls for new technologies, and AI in particular. The European Commission's High Level Expert Group on Artificial Intelligence stated that AI is:

a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation. In particular, AI systems can help

to facilitate the achievement of the UN's Sustainable Development Goals, such as promoting gender balance and tackling climate change, rationalising our use of natural resources, enhancing our health, mobility and production processes (High-Level Expert Group on Artificial Intelligence, 2019).

However, there is a concrete risk in using AI to develop sustainable solutions to the environmental crisis without critically analyzing what they actually mean and adopting a position that we might call “techno-enthusiastic”, i.e., an overly positive, uncritical perspective on the use of technology (Coget, 2017). The manifestation of techno-enthusiasm can culminate in a failure to adequately recognize and evaluate the disruptive effects of AI. These disruptions permeate multiple dimensions, including social, ethical, political, and environmental domains (Floridi, 2022), and can also lead to greenwashing, which is a particular form of ethics-washing that attempts to obscure the environmental impact of a particular AI technology and hide the ecological consequences of certain AI technologies (Heilinger *et al.*, 2023).

All that considered, in addressing the intersection of AI and sustainability, we must first wonder what we mean by sustainable AI. Indeed, we should distinguish between AI that is used to pursue sustainability goals and AI that is implemented in a sustainable manner. This means that, on the one hand, AI can be seen as a promising tool to address the devastating consequences of the current environmental crisis, for example, by allowing to find smart strategies to mitigate the changes in the world climate and its effects on human and other life. On the other hand, research must also address the sustainability of AI technologies by considering energy consumption, the cost of disposing of materials, and the environmental footprint (Van Wynsberghe, 2021). Apart from this distinction, it should also be noted that the sustainability debate often considers only the environmental aspect of the issue, while neglecting the other two pillars of sustainability, namely society and economy (Purvis *et al.*, 2019). All three perspectives should be considered for critical evaluation when incorporating AI into sustainability efforts, in order to establish a scale of values and consider fair priorities by avoiding careless attitudes about the alleged benefits of these new technologies.

In order to evaluate the issues raised previously, this paper is organised as follows: section 2 focuses on the current applications of AI for

sustainability, by introducing some machine learning models and robots used in this field. Section 3 addresses the critical aspects of these tools, which themselves pose a threat to sustainability. Finally, section 4 focuses on one specific example of sustainable AI, i.e., Bayesian networks-based models, and lastly some conclusions are drawn.

Sustainable AI

As Aimee van Wynsberghe notes, sustainable AI should be “a movement to foster change in the entire lifecycle of AI products towards greater ecological integrity and social justice” (van Wynsberghe, 2021). From this perspective, there are myriad AI applications in the area of sustainability, especially following the release of the 17 Sustainable Development Goals (SDGs) set by the United Nations General Assembly in 2015 (*Transforming our World: The Sustainable Development Agenda to 2030*), which encompass environmental, social, political and economic concerns. This gave rise in 2017 to the AI for Good initiative, a digital platform of the United Nations’ International Telecommunication Union where AI innovators and problem solvers learn, discuss and network to find practical AI solutions to advance the SDGs. In addition, Luciano Floridi’s work on the ethics of AI in relation to sustainability is crucial in this area (Mazzi & Floridi, 2023). Behind the idea of “AI for social good” lies of course the hope that the development and use of AI may positively impact individuals, society and environments (Floridi, 2019). There are many applications of AI for social good, that encompass all sectors (Floridi *et al.*, 2020), but on the other hand, it is well-known that AI can also be overused or used for unethical purposes (King *et al.*, 2020). Hence, an ethical analysis is a crucial element of AI for social good, and it is promoted by different stakeholders – from private companies to governments of countries that are including AI in their national strategies. Above all, the common goal that should guide AI for social good is the so-called principle of beneficence, which states that AI should benefit people and the natural world, and hence, should aim to deliver environmentally and socially sustainable outcomes (Mazzi & Floridi, 2023).

Ricardo Vinuesa and colleagues have presented and widely discussed the implications of how AI can either enable or inhibit the delivery of all 17 SDGs recognized by the 2030 Agenda (Vinuesa *et al.*, 2020). They

begin by defining what is meant by AI, since an unambiguous definition is still lacking in literature:

We consider as AI any software technology with at least one of the following capabilities: perception – including audio, visual, textual and tactile (e.g., face recognition), decision-making (e.g., medical diagnosis systems), prediction (e.g., weather forecast), automatic knowledge extraction and pattern recognition from data (e.g., discovery of fake news circles in social media), interactive communication (e.g., social robots or chat bots), and logical reasoning (e.g., theory development from premises). This view encompasses a large variety of subfields, including machine learning (Vinuesa *et al.*, 2020, p. 1).

Concerning machine learning models for SDGs, for instance, hybrid models based on the support vector regression (SVR) and particle swarm optimization (PSO) can be used to predict precision energy usage from supplied data, and also some kinds of predictive algorithms based on recurrent neural networks are useful to make future predictions about climate change. Moreover, generative adversarial networks (GANs) that can learn deep representations without a large amount of annotated training data are widely used in smart cities for generalization capabilities, and smart transportation modeling and management (Mazzi *et al.*, 2023). Concerning social issues, other AI applications deal with SDG 1, i.e., reducing poverty, such as a machine learning classification model to help track and monitor target poverty in some specific countries (Alsharkawi *et al.*, 2021). It's also worth mentioning the efforts for addressing the issue of the lack of water (SDG 6), by implementing machine learning models to predict the water pressure in some areas of the planet, to optimize the work of pumping to avoid waste and damage that would affect the local communities (Pedemonte, 2020), and also for decision-making task in for water management (Bromley, 2005; Phan *et al.*, 2016). Concerning decision-making processes, the usage of Bayesian networks-based algorithms is very common: again, to address SDG 6, David Requejo-Castro (2021) has proposed a data-driven Bayesian network approach: his work combines expert opinion and quantitative data to support informed decision-making. Specifically, he uses Bayesian networks-based structure learning algorithms to replicate composite indicators-based conceptual frameworks, which represent experts' knowledge and identifies interlink-

ages associated with a complex context, coupled with bootstrapping, to reduce results uncertainty, and with a comprehensive result robustness analysis. His results, validated on SDG 6, show that this combined approach improves model inference capacity, identifies the interlinkages among the variables considered, and can be useful for analysis of the complexities in different contexts. Concerning Bayesian models, it is worth also mentioning the work of Juhwan Kim and his team, who developed a statistical method that combines social network analysis and Bayesian modeling, to provide a technological hierarchical structure that can be applied to understand and measure the sustainability of AI technology. In this case, Kim and his team used as a benchmark the International Patent Classification codes of patent documents related to sustainable AI technology, and the final evaluation of the results was assigned to the domain experts (Kim *et al.*, 2018).

For what concerns robotics, it is worth underlining the usage of robots in the field of sustainable development. Specifically, an emerging research direction is social robotics for sustainability, which can be defined as the application of social robotics technologies, exploiting multimodal communication modes based on social cues (e.g., emotions, body language) to promote sustainable behavior among users and contribute to the achievement of the SDGs (Alfieri *et al.*, 2022). Since most of the research about sustainability is focused on changing humans' habits and behaviours, the mainstream tendency within robotics is related to the use of Persuasive Social Robots. A persuasive social robot is an embodied agent that can interact socially with humans and significantly influence or change their behavior, attitudes or cognitive processes (Siegel *et al.*, 2009). On this path, recent studies have shown how social robots may be used as persuasive agents to encourage people to behave sustainably (Beheshtian *et al.*, 2020). Indeed, these robots can influence users' awareness of environmental resources: for instance, they can have an impact on reducing energy consumption, e.g., by congratulating or disapproving the users concerning their behaviour during the setting of a washing machine (Ham & Midden, 2014). Moreover, social robots can contribute to the improvement of waste sorting by children, such as the robot Pepper which by playing in a competition with kids encourages them to be more aware of materials' recycling (Castellano *et al.*, 2021). Additionally, Pepper is useful also to enhance sustainable behaviour in shared living spaces (Beheshtian *et al.*, 2020).

The aforementioned examples are of course non-exhaustive of all AI applications in the field of sustainability. They have been introduced here to give the reader an idea of how broad the field of AI for sustainability is, and that a critical approach must take into account all the different facets of not only the concept of AI but also of sustainability itself. Moreover, it is crucial not to forget that all the benefits derived from AI come with a very high price. The next section will discuss some critical aspects of these tools, which themselves pose a threat to sustainability.

The dark side of AI

As Kate Crawford (2021) stresses, the term AI evokes the ideas of algorithms, mysterious computing processes and vague software scattered in the cloud, but in the very first place, nothing of such things might work without a very concrete component: lithium. In the USA, the vast majority of lithium comes from the Nevada valleys, and its extraction processes are at the center of protests by environmentalists: “the mining that makes AI is both literal and metaphorical. The new extractivism of data mining also encompasses and propels the old extractivism of data mining” (Crawford, 2021, 31). Mining brings with it a history of exploitation of land and people, and it is just the very first of AI’s unsustainable features. After the lithium is extracted and refined, it goes to constitute the components of all AI computing processes. As is known, the processes need great quantities of electricity, which is largely provided by fossil-fuel-based energy supply. Hence, this is associated with greenhouse gas emissions, which of course exacerbate climate change. AI also puts a strain on energy grids, which can lead to power shortages. This is particularly problematic in regions with limited energy resources, as the energy requirements of AI systems can in some cases necessitate the construction of new power plants, which in turn contributes to environmental degradation (Crawford, 2021).

The critical issues of AI from an environmental point of view are those that emerge most intuitively, but the problems related to the economy and society should not be forgotten either. Vinuesa and colleagues, in their analysis of AI’s impact on SDGs, conclude that current technologies, if used at scale, can contribute to the achievement of 90% of environmental goals, 70% of economic goals and 82% of social goals (Vinuesa *et al.*, 2020). These claims, however, should be scrutinized carefully. Indeed, for

instance, Vinuesa and colleagues do not specify the risk of developing a type of AI that targets one SGD at the expense of another: “if an AI application, e.g., for analysing strategies for protecting biodiversity, recommends the abandonment of long-inhabited, culturally relevant settlements (and thus calls for displacement of some peoples), the net effect on sustainability on earth may be zero or negative” (Heilinger *et al.*, 2023, 3). Of course, not every attempt at one SDG implies the violation of another one, but the competing interests among the different targets should be taken into account when implementing AI addressed to singular sustainability goals. Something that should not be overlooked, too, as stressed by Crawford (2021) and Mark Coeckelbergh (2021) is that using AI for sustainability may also lead to power concentration, and ordering all human activities under the requirement of a climate adaptation or mitigation opens the possibility for a violation of legitimate freedoms. This gains an even more concerning dimension if we consider that AI itself might not be as powerful to achieve the climate goals as often presented, especially if it generates a false sense of sufficient engagement perceived by governments and communities (Heilinger *et al.*, 2023).

Above all, what is even more concerning and should be specifically avoided in the debate and the research about AI and sustainability, is the idea that we may find enough applications that optimize or compensate for all the unsustainable processes or habits performed by humankind. This techno-enthusiasm can result in a techno-solutionism mindset, which is particularly dangerous specifically when it comes to address social issues (Mozorov, 2013). Indeed, attempts to make society more sustainable through technology presuppose that many current social injustices can be easily fixed just by automation. This assumption completely flattens out the complex dynamics underlying social injustices, which in most cases require political and social interventions. These interventions sometimes have little if anything, to gain from technological assistance (Heilinger *et al.*, 2023). As previously anticipated, an unsustainable social cruciality is represented by the business model some AI companies are based on: from labour exploitation to questionable data-sourcing strategies (Chan *et al.*, 2021). Data are regarded as the new oil, and this means that the power a data-collecting company gains over those it collects data from, is further contributing to socially unsustainable developments.

All that considered, the central point of the discussion about AI and sustainability is that we should take into account all the different facets

of the topic, which are often neglected by the literature: sustainability does not relate to the environment only, and technology is not always the panacea for all evils. On this ground, Jan-Christoph Heilinger and colleagues have proposed a comprehensive evaluation of any AI-based technology that attends to two different and complementary dimensions: an assessment according to the ends and outcomes of AI, i.e., considering the actual effects of using AI to advance environmental or social sustainability, and an assessment according to the means, i.e., considering the environmental and the social sustainability of AI as a tool itself (Heilinger *et al.*, 2023). Heilinger and colleagues' conclusions are relevant in the debate since they propose a "terminological distinction between thin and thick accounts of sustainability" (Heilinger *et al.*, 2023, p. 9): only if the two dimensions described above – which mutually involve the social and the environmental aspect of sustainability by considering them both from the means and the perspective of the end – an AI can be labeled as "thick" sustainable. In all other cases, it can be at best justified as "thin" sustainable: calling sustainable something actually not-sustainable is of course a distortion of reality, which leads to ethics-washing. Despite no AI-based technology should, as yet, qualified as sustainable under these assumptions, such a strong distinction is needed, since "sustainability is a powerful notion and should be kept as an ambitious goal, not be cheapened to accommodate, e.g., the economic or political interests [...] at the expense of environmental and social concerns" (Heilinger *et al.*, 2023, p. 9).

Hence, the take-away message of these two sections is to make the reader aware of all the possible declinations of both sustainable and unsustainable AI and to give her a glimpse of the rich and sensitive debate surrounding these notions. On these grounds, in the next section, the paper focuses on some possible strategies to make AI more sustainable – by considering "sustainable" not only in the mainstream sense of less environmental impact – and taking as a specific relevant stance a type of algorithm used for decision-making tasks.

Paths to be followed: a representative case

There is plenty of literature on how to make AI more sustainable (van Wynsberghe, 2021; Anthony *et al.*, 2020), which considers all three pil-

lars of sustainability, i.e., environment, economy and society (Purvis *et al.*, 2019). As can be easily imagined, the most obvious strategies to make AI more efficient sustainable terms of sustainability fall into the realm of environmental policy. They range from developing algorithms that require less computer power to train and run, to optimising the hardware used in AI systems, to using renewable energy sources to run AI systems.

When it comes to algorithmic processes, we have two ways of defining their sustainability: on the one hand, we can consider “sustainable” algorithmic processes that deliver the fastest results with the least resource consumption, i.e., environmentally sustainable in terms of computing power; on the other hand, those that are socially sustainable, i.e. transparent and explainable, not only to the developers themselves, but also to the community. A representative example of a type of algorithmic process that could be considered sustainable – though not in the traditional sense of the term “ecological” – is Bayesian network-based models, used for decision-making tasks. As we saw in section 2, there are several examples of these types of models being used in the field of sustainability, as they are used to meet some of the SDGs or to assess the environmental impact of AI tools themselves. In addition, they have been used in robotics by Pablo Lanillos and Gordon Cheng (2018), who implemented a computational model that enables a robot to infer its own body configuration by relying it on predictive coding, which could be considered a sustainable framework. By predictive coding, or predictive processing, is meant a hierarchical model which is considered a unified account for perception, action and cognition. According to this view, the brain is a predictive machine that tries to predict its next states on the basis of information from previous interaction with the environment (Friston *et al.*, 2011; Clark, 2023). This means that it always tries to minimise the probability of prediction errors and avoid states with a high degree of surprise, which corresponds exactly to the concept of Bayesian inference. In the context of Lanillos and Cheng’s model, predictive processing is used for a computational perceptual model that allows any multisensory humanoid robot to learn, infer and update its body configuration. This model enables generic multisensory integration by integrating different sources of information (tactile, visual and proprioceptive): the robot estimates and adapts its body configuration based only on sensory information.

Predictive processing comes with the free energy principle (FEP), which is to be understood as a general methodology for resource opti-

misation. In its original formulation, FEP is a mathematical formula that states that agents that exist do so because they can survive and maintain their equilibrium by minimising free energy. This minimisation of energy is achieved through active inference, i.e., changes in state and suppression of prediction errors through the information derived from the history of previous interactions with the environment. Active inference is appealing for robotic applications where the dynamics of the robot or the task are uncertain: for estimation, adaptive control, fault-tolerant control, prospective planning and complex cognition skills (human-robot collaboration, self/other distinction). Prior to Lanillos and Cheng, Jun Tani and colleagues (2003) and Leo Pio Lopez and colleagues (Pio Lopez *et al.*, 2016) studied models similar to active inference for real robots. Similar approaches have also been used for industrial manipulators (Pezzato *et al.*, 2020), and for implementing active vision in robots in simulation environments (Van de Maele *et al.*, 2021).

All the generative models based on Bayesian-networks described above, based on FEP and active inference, could be considered in some sense more sustainable than more complex models (e.g. neural models based on feedforward architectures) in terms of fairness and explainability (Imai & Jiang, 2022), since Bayesian networks resemble explicit reasoning, albeit probabilistic. Indeed, they track the propagation of events through multiple ambiguous points, where the event diverges probabilistically between paths, and at any point in the network, the probability of that node being visited depends on the joint probability of the preceding nodes: this means that all processes can be tracked and understood. They also refer to FEP, which provides a very general framework for resource optimisation. On the other hand, this type of model relies on a basic decision theory that assumes an optimal decision maker who is able to calculate and choose the move that maximises the utility function at each stage of problem solving. These theories of maximising expected utility have been criticised as being computationally intractable (Kwisthout & van Rooij, 2020; Lieto, 2021), since they are computationally demanding for systems with a large number of random variables, and therefore lead to an exponential increase in computational effort – and energy consumption – as the number of states increases.

Taking all this into account, Bayesian networks-based algorithms could be considered sustainable in some ways. On the one hand, they are socially sustainable, since they are transparent and model processes

that can be easily traced from start to finish, and on the other hand, they are sustainable in a broader sense in that they relate to a general framework of resource optimisation than can be applied to various kinds of organisations, from biological organisms to complex systems. Of course, they are also unsustainable from other points of view, such as the waste of energy in calculation, and therefore they can be seen as an interesting example of how multi-layered and complex the question of the concept of sustainability is.

Conclusion

This paper has outlined the problematic nature of the debate on AI and sustainability and emphasised that we need to be careful in harnessing the potential of AI to facilitate sustainable solutions to avoid the dangers of a technological solution mentality. Indeed, given the uncontrolled explosion of literature on AI and sustainability, finding the right way to avoid both techno-enthusiasm and distrust of technology is not trivial. The example of Bayesian network-based models for decision-making tasks was used to show how problematic it is to define the sustainability of an AI tool. It is also crucial to consider that “possibly, genuine sustainability will ultimately even remain out of reach when exclusively focussing on the technology itself, without implementing significant reforms in the wider social and political environment within which AI- based technologies are developed, produced, sold, and used” (Heilinger *et al.*, 2023, p. 9). Therefore, a comprehensive and critical understanding of the implications of the use of AI remains necessary. This can circumvent potential disruptions and ethical issues and pave the way for truly sustainable progress.

References

- Alfieri, I., Fleres, A., Damiano, L. (2022). Workshop Eco-socio-botics 2022 – Social Robotics for Sustainability. *14th International Conference ICSR 2022*, Florence, Italy, December 13–16, 2022.
- Alsharkawi, A. *et al.* (2021). Poverty Classification Using Machine Learning: The Case of Jordan. *Sustainability*, 13 (3), 1412. <https://doi.org/10.3390/su130314121>
- Anthony, L.F.W., Kanding, B., Selvan, R. (2020). *Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models*. ArXiv:2007.03051
- Beheshtian, N. *et al.* (2020). GreenLife: A Persuasive Social Robot to Enhance the Sustainable Behavior in Shared Living Spaces. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (NordiCHI '20), October 25-29, Tallinn, Estonia, ACM, New York, NY, USA.
- Bromley, J. (2005). *Guidelines for the Use of Bayesian Networks as a Participatory Tool for Water Resource*. Wallford.
- Castellano, G. *et al.* (2021). PeppeRecycle: Improving Children's Attitude Toward Recycling by Playing with a Social Robot. *International Journal of Social Robotics*, 13, 97-111. <https://doi.org/10.1007/s12369-021-00754-0>
- Chan, A., Okolo, C., Terner, Z., Wang, A. (2021). *The Limits of Global Inclusion in AI Development*. <https://arxiv.org/pdf/2102.01265>
- Clark, A. (2023). *The Experience Machine: How Our Mind Predict and Shape Reality*. Allen Lane.
- Coeckelbergh, M. (2021). AI for Climate: Freedom, Justice, and Other Ethical and Political Challenges. *AI and Ethics*, 1(1), pp. 67-72. <https://doi.org/10.1007/s43681-020-00007-2>
- Coget, J. (2017). Technophobe vs. Techno-enthusiast: Does the Internet Help or Hinder the the Balance Between Work and Home Life? *Academy of Management Perspectives*, 25, 1. <https://doi.org/10.5465/amp.25.1.95>
- Crawford, K. (2021). *Atlas of AI. Power, Politics and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Floridi, L. (2019). What the Near Future of Artificial Intelligence Could Be. *Philosophy & Technology*, 32(1), 1-15. <https://doi.org/10.1007/s13347-019-00345-y>
- Floridi, L. (2020). The Green and the Blue: A New Political Ontology for a Mature Information Society. SSRN Scholarly Paper ID 3831094. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3831094>
- Floridi, L. (2022). *Etica dell'intelligenza artificiale: Sviluppi, opportunità, sfide*. Raffaello Cortina Editore.
- Friston, K., Mattout, J., Kilner, J. (2011). Action Understanding and Active Inference. *Biological Cybernetics*, 104(1), 137-160.
- Ham, J., Midden, C. (2014). A Persuasive Robot to Stimulate Energy Conser-

- vation: The Influence of Positive and Negative Social Feedback and Task Similarity on Energy-Consumption Behavior. *International Journal of Social Robotics*, 6(2), 163-171. doi: 10.1007/s12369-013-0205-z
- Heilinger, J., Kemp, H., Nagel, S. (2023). Beware of Sustainable AI! Uses and Abuses of a Worthy Goal. *AI Ethics*. <https://doi.org/10.1007/s43681-023-00259-8>
- High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. B-1049 Brussels.
- Imai, K., Jiang, Z. (2022). Principal Fairness for Human and Algorithmic Decision-Making. *ArXiv*, arXiv:2005.10400 <https://doi.org/10.48550/arXiv.2005.10400>.
- ITU. AI4Good Global Summit. Available online: <https://aiforgood.itu.int>
- Kim, J. *et al.* (2018). Sustainable Technology Analysis of Artificial Intelligence Using Bayesian and Social Network Models. *Sustainability*, 10(1).
- King, T. C., Aggarwal, N., Taddeo, M., and Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26(1), 89-120. <https://doi.org/10.1007/s11948-018-00081-0>
- Kwisthout J., van Rooij, I. (2020). Computational Resource Demands of a Predictive Bayesian Brain. *Comput Brain Behav*, 3, 174-188. <https://doi.org/10.1007/s42113-019-00032-3>
- Lanillos, P., Cheng, G. (2018). Adaptive Robot Body Learning and Estimation through Predictive Coding. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4083-4090.
- Lieto, A. (2021). *Cognitive Design for Artificial Minds*. Routledge.
- Mazzi, F. *et al.* (2023). *AI in Support of the SDGs: Six Recurring Challenges and Related Opportunities Identified Through Use Cases*. In Mazzi, F. & Floridi, L. (Eds.), *The Ethics of Artificial Intelligence for the Sustainable Development Goals* (pp. 9-33). Springer.
- Morozov, E. (2013). *To Save Everything, Click Here. The Folly of Technological Solutionism*. Perseus.
- Pedemonte, V. (2020). *AI for Sustainability: An Overview of AI and the SGDS that Contribute to the European Policy Making*.
- Pezzato, C., Baioumy, M., Corbato, C. H., Hawes, N., Wisse, M., Ferrari, R. (2020). Active Inference for Fault Tolerant Control of Robot Manipulators with Sensory Faults. In *International Workshop on Active Inference* (pp. 20-27). Springer.
- Phan, Th.D. *et al.* (2016). Applications of Bayesian Belief Networks in Water Resource Management: A Systematic Review. *Environmental Modelling and Software*, 85, 98-111. <https://doi.org/10.1016/j.envsoft.2016.08.006>
- Pio-Lopez, L., Nizard, A., Friston, K., Pezzulo, G. (2016). Active Inference and Robot Control: A Case Study. *Journal of The Royal Society Interface*, 13(122).
- Purvis, B., Mao, Y., Robinson, D. (2019). Three Pillars of Sustainability: In Search of Conceptual Origins. *Sustainability Science*, 14, 681-695. <https://doi.org/10.1007/s11625-018-0627-5>

- Requejo Castro, D. (2021). *Data Driven Bayesian Networks Modelling to Support Decision-making: Application to the Context of Sustainable Development Goal 6 on Water and Sanitation* [PhD thesis].
- Saetra, H. (2021). AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System. *Sustainability*, 13, 1738. <https://doi.org/10.3390/su13041738>
- Siegel, M., Breazeal, C., Norton, M. I. (2009). Persuasive Robotics: The Influence of Robot Gender on Human Behavior. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2563-2568. doi: 10.1109/IROS.2009.5354116
- Tani, J. (2003). Learning to Generate Articulated Behavior through the Bottom-up and the Top-down Interaction Processes. *Neural Networks*, 16(1), 11-23.
- United Nations General Assembly (2015). *Transforming our World: The Sustainable Development Agenda to 2030*.
- Van De Maele, T. *et al.* (2021). Active Vision for Robot Manipulators Using the Free Energy Principle. *Frontiers in Neurorobotics*, 15. doi: 10.3389/fnbot.2021.642780
- van Wynsberghe, A. (2021). Sustainable AI: AI for Sustainability and the Sustainability of AI. *AI Ethics*, 1(3), 213-218. <https://doi.org/10.1007/s43681-021-00043-6>
- Vinuesa, R., Azizpour, H., Leite, I. (2020). The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233. <https://doi.org/10.1038/s41467-019-14108-y>