

# Chapter 30

## Responsible Machine Learning for Ethical Artificial Intelligence in Business and Industry

**Deepak Saxena**

 <https://orcid.org/0000-0002-9331-3799>

*Birla Institute of Technology and Science, Pilani, India*

**Markus Lamest**

*Trinity College Dublin, Ireland*

**Veena Bansal**

*Indian Institute of Technology, Kanpur, India*

### ABSTRACT

*Artificial intelligence (AI) systems have become a new reality of modern life. They have become ubiquitous to virtually all socio-economic activities in business and industry. With the extent of AI's influence on our lives, it is an imperative to focus our attention on the ethics of AI. While humans develop their moral and ethical framework via self-awareness and reflection, the current generation of AI lacks these abilities. Drawing from the concept of human-AI hybrid, this chapter offers managerial and developers' action towards responsible machine learning for ethical artificial intelligence. The actions consist of privacy by design, development of explainable AI, identification and removal of inherent biases, and most importantly, using AI as a moral enabler. Application of these action would not only help towards ethical AI; it would also help in supporting moral development of human-AI hybrid.*

### INTRODUCTION

Breaking away from the realm of science fiction, artificial intelligence (AI) systems have become the reality of modern life. Kaplan and Haenlein (2019) define AI as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks

DOI: 10.4018/978-1-7998-6985-6.ch030

through flexible adaptation”. AIs in modern life not only provide recommended course of actions to those working in business and industry, but they are also effectively making decisions for us. In our everyday life, they strongly influence what we watch during our leisure time, who we interact more with on social networks, the videos we watch and the music we listen to, what we buy on e-commerce websites and on what rate, what interest rate and credit limits we are offered on our credit cards, what route our taxis take and so on. The list of AI applications is endless. However, AI models are usually opaque, and one often knows very little on the decisions-making algorithm used by AIs (Burrell, 2016; de Saint Laurent, 2018; McQuillan, 2018; Mackenzie, 2019). Since AIs have so much influence on our lives, it is an imperative to focus our attention on the ethics of AI.

At the same time, AIs are built by humans, often using the process of machine learning. Hence, the choices made by the AI developers (who design the AI) and the managers (who decide to implement the recommendations from the AI) are crucial. This paper offers some suggestions towards responsible machine learning building on the idea of human-AI hybrid (Bansal et al., 2019a, 2019b; Dellermann et al., 2019; Jarrahi, 2018; Peeters et al., 2020). Taking a socio-technical systems perspective (Andras et al., 2018), the term human-AI hybrid indicates that each subsystem (AI and the humans) brings a unique skillset to the table and can help in ethical development of the overall system. In so doing, the chapter takes machine learning as a common denominator of AI and introduces a responsible machine learning framework towards ethical AI. For developers and managers, the chapter suggests actionable insights that can be used in industry and business.

The remainder of the chapter is as follows. The next section provides some background in terms of the notion of artificial intelligence, the machine learning process, and the concept of distributed morality in the context of socio-technical systems. Thereafter, the building blocks of responsible machine learning are discussed with the implications they have for the AI developers and managers. The building blocks are aligned with the classical machine learning stages. Based on the discussion, recommendations are presented for the managers and developers. This is followed by noting some limitations and future development, with last section concluding the chapter.

## **BACKGROUND**

### **The Notion of Artificial Intelligence**

While the notion of AI has been popular from the movies like *The Terminator* or *Ex Machina*, there are two crucial limitations related to the popular perception of AI. The first limitation relates to the notion of intelligence. Strictly speaking, the use of the term intelligence in AI does not relate to the entirety of human intelligence but refers mostly to the logical-mathematical intelligence (Brock, 2018; Vetrò et al., 2019). When developers build an AI, they try to automate and optimize the processing that requires data manipulation and logical reasoning. While there have been efforts to design emotionally intelligent AI (Schuller & Schuller, 2018) or AIs with kinaesthetic intelligence (e.g., self-driving cars), the efforts are yet to see the light of the day in wider society. It is with this caution the term ‘intelligence’ is to be used for AI. Therefore, while AIs are fit to replace humans in the tasks that require an application of logical-mathematical intelligence, human intervention is still needed (Brock, 2018; Ross, 2018; Wilson et al., 2017) due to other forms of intelligence they possess, especially when it comes to ambiguous situations and inductive reasoning. For this reason, this paper draws from the idea of the human-AI hybrid

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/responsible-machine-learning-for-ethical-artificial-intelligence-in-business-and-industry/285003](http://www.igi-global.com/chapter/responsible-machine-learning-for-ethical-artificial-intelligence-in-business-and-industry/285003)

## Related Content

---

### Classical Clustering Technique for Segmentation of Skin Cancer Image

Saravana Kumar V., Kavitha M., Anantha SivaPrakasam S. and Bavya S. (2024). *Applications of Parallel Data Processing for Biomedical Imaging* (pp. 261-275).

[www.irma-international.org/chapter/classical-clustering-technique-for-segmentation-of-skin-cancer-image/345600](http://www.irma-international.org/chapter/classical-clustering-technique-for-segmentation-of-skin-cancer-image/345600)

### Framework and Model for Cryptocurrency Innovation and its Impact on Economic Transformation

Ahmed Ashoor and Kamaljeet Sandhu (2021). *Research Anthology on Blockchain Technology in Business, Healthcare, Education, and Government* (pp. 217-229).

[www.irma-international.org/chapter/framework-and-model-for-cryptocurrency-innovation-and-its-impact-on-economic-transformation/268600](http://www.irma-international.org/chapter/framework-and-model-for-cryptocurrency-innovation-and-its-impact-on-economic-transformation/268600)

### Role of Artificial Intelligence in Healthcare: Technologies, Applications, Challenges, and Issues

S. Chidambaram (2023). *Contemporary Applications of Data Fusion for Advanced Healthcare Informatics* (pp. 412-428).

[www.irma-international.org/chapter/role-of-artificial-intelligence-in-healthcare/327731](http://www.irma-international.org/chapter/role-of-artificial-intelligence-in-healthcare/327731)

### New Era in the Supply Chain Management With Blockchain: A Survey

Jesus Alvarado and Malka N. Halgamuge (2021). *Research Anthology on Blockchain Technology in Business, Healthcare, Education, and Government* (pp. 1770-1794).

[www.irma-international.org/chapter/new-era-in-the-supply-chain-management-with-blockchain/268688](http://www.irma-international.org/chapter/new-era-in-the-supply-chain-management-with-blockchain/268688)

### Big Data and Big Data Analytics: A Review of Tools and its Application

Sam Goundar, Akashdeep Bhardwaj, Shavindar Singh, Mandeep Singh and Gururaj H. L. (2021). *Applications of Big Data in Large- and Small-Scale Systems* (pp. 1-19).

[www.irma-international.org/chapter/big-data-and-big-data-analytics/273917](http://www.irma-international.org/chapter/big-data-and-big-data-analytics/273917)