AI Ethics for Industry 5.0 – From Principles to Practice

Alexandru Constantin Ciobanu [1], and Gabriela Meșniță [1]

[1] *Alexandru Ioan Cuza University of Iasi, Carol Blv. 11, Iași, 700506, Romania*

Abstract

Industry 5.0 brings new challenges to the way humans are organizing in societies that could be sustainable from an economical and environmental perspective. If Industry 4.0 is rather focused on productivity and automation, Industry 5.0 claims to be a human-centric one. One key technical component of both Industry 4.0 and 5.0 is represented by the Artificial Intelligence (AI) systems in their different forms of applicability. Our paper argues that, if Industry 5.0 claims to be a human-centric orientated, then we as human beings must be able to create, implement and control AI systems that are ethical, sustainable, and reliable. For that we are suggesting an AI Ethical framework that aims to offer a perspective of how ethical AI could be practically implemented. Our current paper is based on previous research that we conducted and is still subject of further research developments meant to validate the proposed framework.

Keywords

Artificial intelligence, ethics, Industry 5.0, practical, framework.

1. Introduction

Within the human history, the technical progress changed not only the way different goods were produced, but also how the overall societies are functioning. Technologies have evolved disruptively since their inception, and the literature defined four major industrial revolutions (well known as Industry 1.0 to Industry 4.0).

While Industry 4.0 is still happening and more and more businesses are transitioning to it, we started few years ago to discuss also about Industry 5.0. Plenty of definitions about Industry 5.0 could be found but essentially, Industry 5.0 couldn't be considered as new industrial revolution but rather as a transition to a potential 6th one [1]. Industry 5.0 is introducing a shift from Industry 4.0 (based on a technological–centric approach) to a human-centric approach, where organizations are seeking to obtain more than just business productivity through innovation. Within this new industry context, humans can easily work with robots and Artificial Intelligence (AI) systems, while being productive and more creative. While some research [2, 3] are showing the conceptual benefits of the Industry5.0, others [4, 5] are arguing this is "building a complex and hyperconnected digital networks", where one potential challenge is that businesses are not focused enough on the organizational issues arising from human–computer collaboration. Therefore, a series of ethical concerns needs to be clarified. We understand it is essential to have in place the right governance and applied frameworks that can tailor the way AI technology systems are developed and used to avoid potential unethical challenges.

Most of the current researchers [6-8] have conducted studies where the features of Industry 5.0 are directly linked with the economy and society, in a way where these new features will not only improve productivity, but also these must improve *the quality of life.* They argued on what could be the frameworks and guidelines that can be embedded within the technology development area so that organizations may minimize business uncertainty while boosting innovation that may address current social issues.

Proceedings of the Workshop of I-ESA'22, March 23–24, 2022, Valencia, Spain
EMAIL: alexandru.ciobanu@feaa.uaic.ro (A. Ciobanu); gabriela.mesnita@feaa.uaic.ro (G. Meșniță)
ORCID: 0000-0003-1681-1505 (G. Meșniță)


Authors such as Mavrodieva and Shaw [9] link the Industry 5.0 with what they call "a Super Smart Society, with a focus on individual needs and capabilities. The concept envisages a merge between the real (physical) world with the cyberspace to efficiently collect more precise and personalized data for improved problem solving and value creation". This idea seems already to be present within the latest announcement of Microsoft and Facebook, related to the *Metaverse* [10, 11], as being the platform that could potentially link the digital assets to the real-world economy activities. In other words, the metaverse could be a defining key element of Industry 5.0 where we can embed the real world into the computing so that organizations may achieve more work flexibility for their employees and innovation.

While Industry 5.0 is still depicted from a conceptual point of view, recent research [12] argues on the fact that "a value-oriented and ethical technology engineering in Industry 5.0 is an urgent and sensitive topic". We considered this as being an important point, because it is emphasizing the importance of developing further technologies, while enabling and maintaining human–machine cooperation in a balanced symbiosis from an ethical perspective.

One of the key technical elements of Industry 5.0 is the one referring to *Artificial Intelligence systems*. For our paper we are understanding through AI "an umbrella term to cover a set of complementary techniques that have developed from statistics, computer science and cognitive psychology. While recognizing distinctions between specific technologies and terms (e.g., artificial intelligence vs. machine learning, machine learning vs. deep learning), it is useful to see these technologies as a group, when considering how to support development and use of them" [13].

The European Commission has several priorities in relation to Industry 5.0 [14] including the following one "*adopting a human-centric approach for digital technologies including artificial intelligence"*. Therefore, they have published a Proposal for AI regulation [15].

Given the above presented context, the hypothesis of our paper is that the further developments of AI systems aligned with the Industry 5.0 growth will necessary increase the need of an AI ethical framework that need to be implemented and operationalized based on different variables (i.e.: social, cultural, political, religious or legal).

This paper emphasizes how we can achieve innovation and human-machine cooperation envisaged by Industry 5.0, while guaranteeing a series of digital ethics principles? Our goal is to develop a framework that helps to operationalize the AI ethical principles so that these can be embedded in an efficient and controlled manner within the future technological systems characterizing the Industry 5.0. This paper will follow 2 sections: in the first one we present the considered AI ethical principles and the current implementation challenges as described by other researchers, and within the second par we will present our proposed AI ethical framework.

2. AI ethics from principles to practice – challenges

Some of the key concepts of Industry 5.0 are the ones referring to *innovation, cooperation, human-machine interoperability.* Since one key technical component of Industry 5.0 is represented by AI systems themselves, it's important to further understand how we can further develop AI systems in a harmonized way, towards an efficient human-machine interaction driven by innovation while being guided by ethical principles.

Within the last years, the AI ethics was a subject of maximum interest among academics, policy decision-makers, AI professionals or producers as well as the regular public. Most part of the research argued the need of ethics in AI [16-19] to further avoid potential risk associated with AI (such as inequality, biases, privacy, labor market etc.). Other studies defined principles [20, 21] and depicted frameworks [22, 23] on how ethical principles should be embedded within the AI systems. However, as a consensus across the various research published around AI ethics principles, we can identify a challenge related to the gap between high-level principles and practical techniques that can be implemented to design and develop ethical AI systems.

For this research we have considered the principles defined by European Commission in The Ethics Guidelines for Trustworthy Artificial Intelligence [24]: *Respect for human autonomy*, *Prevention of harm*, *Fairness* and *Explicability*.

There are many AI ethics frameworks that were published either by academic researchers or even by AI producers (i.e., Microsoft, Google etc.). Almost all of them mention requirements that an AI system must have to be considered ethical. These requirements are related to essential values such as privacy, fairness, non-discrimination, transparency or explicability, security, and accountability. In other words, we have the theory well documented, but the way these frameworks could be implemented and operationalized by different organizations is still unclear and raising lot of challenges.

Since various stakeholders could be interacting with an AI system while developing, implementing, or using this, it is important to have a unified approach when it comes to operationalizing AI ethics. In the study of VDE AI Ethics Group [25] the current challenges of implementing AI, are grouped within three main categories: *Context-dependence*, *Socio-technical nature of AI* and *Ease of Use.*

Besides these challenges, based on previous research we have conducted [26], we identified another one related to the overall governance that we can apply on such framework, from a functional perspective. Some of the AI systems are capable to auto adapt or learn from previous processes. Hence it is important that an AI ethical framework (post implementation) should be capable to be always up to date via *feedback channels* that need to exist between developers, consumers, and other stakeholders. Additionally, an AI consumer/producer needs to be able to measure the success of an applied AI ethics framework and improve when needed. This should be a *continuous* process and not a one-time job.

Obviously, not every AI system will present same risks depending on how, where and by whom it's used. EU HLEG [24] initiative identified 4 main category of risks that may apply to an AI system from the moment is produced to its implementation. Hence an AI ethics framework should be able to differentiate between different categories of risks, while suggesting accordingly different actions.

3. A model for implementing AI ethics framework

Our model is based on the VDE AI Ethics group study [25] that proposed a framework that could operationalize the AI ethical principles based on *variables, criteria's, indicators, and observables* (VCIO). The VCIO model is a multidimensional approach that's addressing the AI ethical principles on several layers of both AI producers and consumers. For this paper we would like to go one step further, considering that an AI ethical system within the industry 5.0 should be *reliable,* and hence an AI ethical framework need to be *verifiable*, *measurable* and capable of assuring a live *feedback* loop between consumers and producer, impacting the training model.

One key component of an AI system is the underlying *training model*. This is the phase in the data science development lifecycle where practitioners try to fit the best combination of weights and bias to an algorithm to minimize a loss function over the prediction range [27]. The purpose of a model training is to build the best mathematical representation of the relationship between data features and a target label such as an AI principle.

Our suggested practical AI ethical framework argues on translating each envisaged principle in an AI training model. This way the model can be subject of validation, testing and deployment. Additionally, each training model, needs to be able to get updated on the associated datasets even after the AI system was implemented, for continuous improvement.

Within the Figure 1 below, we see a concept of a *high-level architecture* of our suggested practical AI ethical framework. This is based on previous research we have conducted, and subject of future research where we are looking for ways of validating that. This architecture framework takes in consideration each of the 4 principles identified by the EU HLEG initiative and is split between two layers *AI Embedded Ethics by Design (AI EED)* and *AI Desired State Configuration (AI DSC)* that are linked through API *systems*. Based on these two linked perspectives we are suggesting an end-to-end approach of operationalizing an AI system that's addressing challenges on both AI producers and consumer's side.

We will further describe this concept and how the two layers (AI EED and AIDSC) can be managed within the framwork. Additionally we will explain the interoperability of these two layers and the role of the API systems.

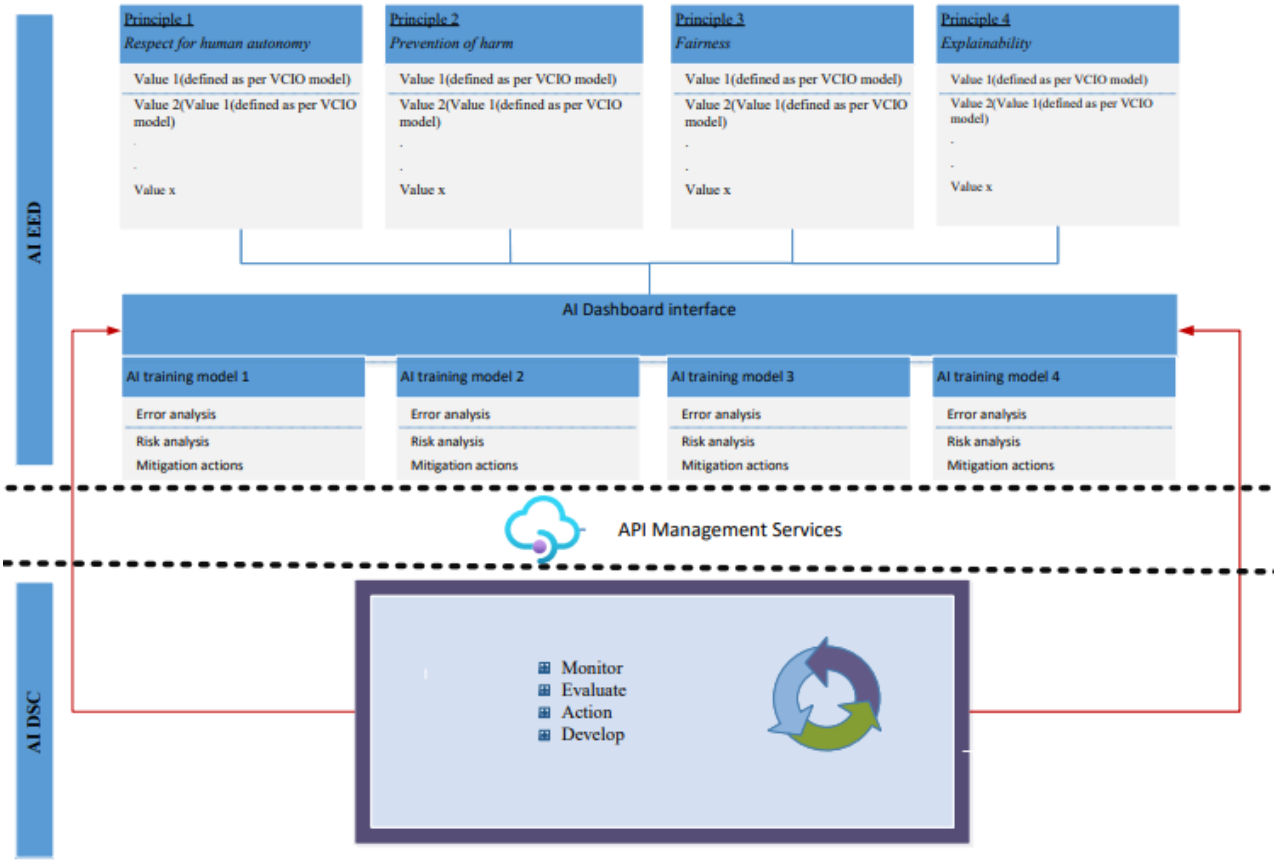

Figure 1: High-level architecture of a practical AI ethical framework

AI Embedded Ethics by Design (AI EED)

On the EED part, each Developer or Data scientist can train his model and the data sets in order to address the required ethical principles that AI solution need to be aware of. The proposed process underlaying the AI EED layer is:

- AI producers will always have an AI Dashboard interface that can be used for the training model.
- Each developer or data scientis from both the AI Producer or the Consumer side can define the AI principles as per the above-mentioned VCIO model.
- Within the AI dashboard interface the developers can create training models for each envisaged principle, instructing the AI system how to aknowledge that.
- Each training model could be tested before being released into production. During the testing there will be considered the potential error, the risks analysis and the actions that could be taken to mitigate that risk and resolve a potential error.

The AI EED layer is considered to be the space that each organization can useit in a *proactive* way before releasing in production a specific AI system, keeping in consideration the cultural context, the industry where that AI system will be deployed and the potential impact as well as who are the other stackholder involved in further managing that system once implemented.

AI Desired State Configuration (AI DSC)

On the other side the AI DSC is way of actively managing the AI system post implementation assuring that the underlying training models are always reliable in different contexts. Considering the characterisitcs of the Industry 5.0 we strongly believe that, any AI system, post migration, need to be managed by humans in a way it can be *monitored*, *measured* on specific criterias, *actioned* and further *developed*. This should also allow non-technical stackholders from the consumer side (such as security officers, data analysts etc.) to continuosly assess the system and provide feedback back on

the AI dashboard interface where the training model can always be re-tested and re-adapted based on the received feedback.

The technical comunication between the above two layers is assured by an API Management System. The feedback (containing structured or un-structured data sets) can be sent post implementation to the AI dashboard interface so that the data scientist and the developers will be further able to ingest that into the training models and this can be a continuos process. This way we can assure an end-to-end approach where the AI system will always be adapted to the contextual reality where they are implemented, allowing the humans to be in control of the AI decisions while focusing on being more creative and bringing more innovation.within their fields of activity.

4. Conclusions

As observed, Industry 5.0 claims to shift from technical productivity systems to a more human-centric approach. In this regard, socio-technical interoperability between humans and AI systems is a mandatory step. The world is intensively digitized, and the new modern workplace is not only the physical one but rather the virtual one. If people need to be more creative while focusing on innovation, we need to build and implement technologies and AI systems that can be guaranteed as ethical and reliable. This can be achieved only if we as human will apply the right governance on the technology while considering the productivity but also a sustainable economy and environment. Our research emphasizes the fact that we need AI ethical frameworks that needs to be practically implemented in a unified approach that's considering the different layers an AI system have before, during and post implementation, as well the different stakeholders that will interact with these technologies.

Our proposed framework aims to mix the theoretical landscape of defining ethical principles with the technical aspects that are underlaying an AI system via two layers approach (AI EED and AI DSC). Additionally, we are suggesting a way (via an API system) where feedback could be collected after the AI system was implemented, and integrated on the initial training model, in a continuous process.

Our paper aimed to present a high-level overview architecture and as a next step, we are looking to develop each component of the suggest framework (AI EED, AI DSC and the interconnecting API) based on different AI systems and their applicability and risk categorization. Additionally, we're looking forward to a way where we can validate this framework in a real live scenario.

5. Acknowledgements

This paper is partially supported by the Competitiveness Operational Programme Romania under project number SMIS 124759 - RaaS-IS (Research as a Service Iasi), POC/398/1/1, SMIS 124759, ctr. nr. 236/21.04.202.

6. References

[1] M. Di Nardo, H. Yu, Special Issue "Industry 5.0: The Prelude to the Sixth Industrial Revolution", Applied System Innovation 4 (2021) 45. doi: 10.3390/asi4030045.

[2] X. Xu, Y. Lu, B. Vogel-Heuser, L. Wang, Industry 4.0 and Industry 5.0 - Inception, conception and perception, Journal of Manufacturing Systems 62 (2021) 530-535. doi: 10.1016/j.jmsy.2021.10.006.

[3] S. Nahavandi, Industry 5.0 - A Human-Centric Solution, Sustainability 11 (2019) 16. doi: 10.3390/su11164371.

[4] V. Özdemir, N. Hekim, Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, "The Internet of Things" and Next-Generation Technology Policy, Omics, 22 (2018) 65-76. doi: 10.1089/omi.2017.0194.

[5] K. Demir, G. Döven, B. Sezen, Industry 5.0 and Human-Robot Co-working, Procedia Computer Science 158 (2019) 688-695. doi: 10.1016/j.procs.2019.09.104.

[6] Y. Zengin, S. Naktiyok, E. Kaygin, O. Kavak, E. Topcuoglu, An Investigation upon Industry 4.0 and Society 5.0 within the Context of Sustainable Development Goals, Sustainability 13 (2021) 2682. doi: 10.3390/su13052682.
[7] B. Aquilani, M. Piccarozzi, T. Abbate, A. Codini, The Role of Open Innovation and Value Co-creation in the Challenging Transition from Industry 4.0 to Society 5.0: Toward a Theoretical Framework, Sustainability 12 (2020) 8943. doi: 10.3390/su12218943.
[8] O. A. ElFar, C.-K. Chang, H. Y. Leong, A. P. Peter, K. W. Chew, P. L. Show, Prospects of Industry 5.0 in algae: Customization of production and new advance technology for clean bioenergy generation, Energy Conversion and Management: X 10 (2021) 100048. doi: 10.1016/j.ecmx.2020.100048.
[9] A. V. Mavrodieva, R. Shaw, Disaster and Climate Change Issues in Japan's Society 5.0 - A Discussion, Sustainability 12 (2020) 1893. doi: 10.3390/su12051893.
[10] S. B. Hall, C. Li, What is the metaverse? And why should we care?, 2021. URL: https://www.weforum.org/agenda/2021/10/facebook-meta-what-is-the-metaverse/.
[11] S. Nadella, Microsoft CEO's on Metavers and Felxible Work, 2021. URL: https://www.youtube.com/watch?v=FRztx0wVuPA .
[12] F. Longo, A. Padovano, S. Umbrello, Value-Oriented and Ethical Technology Engineering in Industry 5.0: A Human-Centric Perspective for the Design of the Factory of the Future, Applied Sciences 10 (2020) 4182. doi: 10.3390/app10124182.
[13] R. Procter, B. Glover, E. Jones, Research in the age of automation, Research 4.0 DEMOS, 2020. URL: https://demos.co.uk/wp-content/uploads/2020/09/Research-4.0-Report.pdf.
[14] European Commission, Industry 5.0, 2020. URL: https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en.
[15] European Commission, 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.
[16] E. Kazim, A. S. Koshyiama, A high-level overview of AI ethics, Patterns, 2 (2021) 100314. doi: 10.1016/j.patter.2021.100314.
[17] L. Floridi, Establishing the rules for building trustworthy AI, Nature Machine Intelligence 1 (2019) 261-262. doi: 10.1038/s42256-019-0055-y.
[18] T. Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines 30 (2020) 99-120. doi: 10.1007/s11023-020-09517-8.
[19] A. Resseguier, R. Rodrigues, AI ethics should not remain toothless! A call to bring back the teeth of ethics, Big Data & Society 7 (2020). doi:10.1177/2053951720942541.
[20] A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines, Nature Machine Intelligence 1 (2019) 389-399. doi: 10.1038/s42256-019-0088-2.
[21] J. L. Zhou, F. Chem. A. Berry, M. Reed, S. J. Zhang, S. Savage, A Survey on Ethical Principles of AI and Implementations, 2020. URL:https://opus.lib.uts.edu.au/bitstream/10453/146673/2/IEEE_ETHAI2020_ethicalAI_Survey.pdf.
[22] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, … E. Vayena, AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Minds and Machines 28 (2018) 689-707. doi: 10.1007/s11023-018-9482-5.
[23] A. Mona, R. Madan, U. Sivarajah, Ethical framework for Artificial Intelligence and Digital technologies, International Journal of Information Management 62 (2022) 102433. doi: 10.1016/j.ijinfomgt.2021.102433.
[24] European Commission, The Ethics Guidelines for Trustworthy Artificial Intelligence (AI), 2021. URL: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html
[25] VDE BertelsmannStiftung, From Principles to Practice: An interdisciplinary framework to operationalise AI ethics, 2019. URL: https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf
[26] A. C. Ciobanu, G. Meşniţă, AI Ethics in Business-A bibliometric approach, REBS 14 (2021) 169-202. doi: 10.47743/rebs-2021-2-0009.
[27] C3.ai, Glossary, 2022. URL: https://c3.ai/glossary/data-science/model-training/.