Human Aspects of Advanced Manufacturing, Production Management and Process Control, Vol. 146, 2024, 31–39
https://doi.org/10.54941/ahfe1005148

Designing Hybrid Intelligence: Understanding the Impact of Human Decision-Making on AI

Melanie Kessler[1], Oliver Antons[2], and Julia Arlinghaus[2]

[1]IFF Magdeburg, Germany
[2]Otto-von-Guericke University, Magdeburg, Germany

ABSTRACT

In many domains such as management, production and government, established control approaches struggle to address increasing complexity in a timely manner, resulting in a demand for more agile methods. Hybrid intelligence and decision support systems are useful approaches to augment human decision-making through artificial intelligence (AI). Various application of AI methods to estimate production parameters or to provide forecasts are discussed in the literature or already being implemented, however, human decision-making is still required for either deciding whether to follow specific suggestions or for monitoring their respective implementation. But human behavioral research has shown that human decision-making is rather biased than fully rational, leading to unintended consequences in the collaborative work of humans and machines. Subsequently, the research stream of hybrid intelligence has gained interest recently, aiming to study the collaboration between humans and machines. We contribute to this issue by combining a systematic literature review on AI and cognitive biases combined with practical insights from discussions with experts in order to derive first guidelines addressing the human factor in the design of AI-based decision support systems for complex production environments.

Keywords: AI, Human decision-making, Hybrid intelligence

INTRODUCTION

Increasing digitalization has provided a new levels of data availability, accuracy and topicality, setting up AI for broad implementation in practice. For human workers, repetitive tasks can be particularity challenging, while AI, decision support systems (DSS) and business intelligence systems have been introduced widely in production environments to alleviate this burden (Chui et al., 2021). Although AI enables smart manufacturing trough the takeover of various monitoring and controlling tasks in the production environment, human decision-making is still required. Especially creative tasks, planning and expertise in decision-making require human involvement and cannot be solved by AI thus far (Kamar, 2016). Based on this need for collaboration between humans and machines the research field of hybrid intelligence has recently gained an impetus. However, human decision-making is often biased leading to a systematic deviation from a rational optimum (Tversky

& Kahneman, 1974). As AI technology requires huge training data sets which are curated based often on previous human decision making, any biases in these data sets are transferred to the AI implementation, subsequently leading to an likewise biased AI decision algorithm. In 2019 researchers found out that an AI algorithm used in US hospitals favored white patients over black patients (Real-Life Examples of Discriminating Artificial Intelligence | by Terence Shin, MSc, MBA | Towards Data Science, 2019). The underlying reason for thus phenomenon of discrimination was found to be due to correlation of healthcare costs and care costs within historic patients data sets. Another example is a hiring algorithm of Amazon, which was biased towards woman over men. This happened because the algorithm used the number of resumes of applicants over the previous years which were mainly from men. Moreover, research has also observed other effect of biases in the collaboration between humans and machines. Phenomena such as the Lead-time-syndrome which describes the effect that human update system-defined lead-times, have been researched intensely in the field of production planning which result from a systematic deviation of humans from a predefined optimum of an AI regarding planned lead times (Bendul & Knollman, 2016).

Especially in complex production environments, the use of AI technology has increased during the last years. Nevertheless, these examples demonstrate the need for an adequate collaboration between humans and machines to fully exploit the associated potentials. Based on a workshop of the authors and further results of discussion with six experts, we shed light on the effects of cognitive biases for AI and humans and derive first propositions for the designing of AI considering these decision-making constraints.

The remainder of this article is structured as follows. First, we combine the literature streams of complex production environments, AI and human behavior. Second, we outline the applied research methodology. Third, we present our findings on cognitive biases and AI in complex production systems. We structure our findings by developing a framework showcasing the relation between human decision-making and AI, as well as the influence of cognitive biases thereupon, deriving first recommendations for designing better hybrid intelligence systems.

AI IN COMPLEX PRODUCTION ENVIRONMENTS IN TIMES OF DIGITALIZATION

Increasing digitalization and the trend towards individual customer products lead to an increasing complexity level in production (Arlinghaus and Antons, 2022). A huge number of possible variants with simultaneously shorter product life cycles and decreasing lot sizes require huge control and monitoring efforts in the production process (Windt et al., 2008). Established control approaches struggle to address this increased complexity in a timely manner, leading to a demand for agile, flexible and adaptive methods (Antons and Arlinghaus, 2022). Therefore, especially complex production environments require support in decision-making by the use of AI technology (Management and Applications of Complex Systems, n.d.). AI is characterized by the "ability of computers to perform cognitive functions associated with human

minds, such as perceiving, reasoning, learning, and problem solving" (Chui, Kamalnath, McCarthy, 2021). In a survey of (Gartner Top 10 Strategic Technology Trends, 2018) regarding technological trends, AI was named as the most strategic technology. Especially in various planning and control tasks in the production environment such as the monitoring of missing parts, determination of inventory level or supplier evaluation the use of AI technology enables fast information gathering and processing. Nevertheless, even leading experts such as Bill Gates and Stephan Hawking predicted the complete replacement of humans trough AI time has shown that still human involvement in decision-making is required (Microsoft's Bill Gates Insists AI Is a Threat - BBC News, 2018). Especially in uncertain and unstructured environments, AI is very limited to solve these decisions adequately and human involvement is still required (Büttner et al., 2022). Further, for the development of the algorithms huge training data and therefore also human support and control of the results are necessary (Burggräf et al., 2018). To fully exploit the potential AI technology can offer the acceptance of the end user is necessary and should therefore be taken into consideration in the designing of an AI technology (Duan et al., 2019).

Therefore, defining the optimum collaboration between humans and AI technology is a key question for designing the future working environment and using the full potential of hybrid intelligence.

Cognitive Biases

Contrary to the common assumption of a rational human behavior research showed that human decision-making is bounded rational and introduced the term of cognitive bias (Kahneman & Tversky, 1979).

Especially when humans are confronted with uncertainty and time pressure in complex environments, they systematically deviate from a rational optimum. Thus, human decision-making is often not completely rational, but distorted by ways of their experiences and personal attitude. A prominent model in cognitive bias research associates this phenomenon to two different areas in the human brain which are involved in the decision-making process, and named them as System I and System II (Kahneman & Tversky, 1979). Whereas system I creates spontaneous impressions, reacts fast and emotional system II is effortful, logical, and controlled (Stanovich & West, 2000).

In the discussion of complementary work of humans and AI it has been shown that humans are good in creative, flexible and empathic decision-making. However, based on the bounded rationality of humans AI technology especially convinces through the ability of fast processing of huge data amounts and recognizing complex patterns therein this data.

In literature, a huge variety of different cognitive bias effects have been researched and discussed intensely. Also, various classification models have been proposed in literature to categorize this plenty of different effects (see for instance (Arnott, 2006; Mehrabi et al., 2019)). Three defined categories of bias effects which occur in AI technology (Mehrabi et al., 2019).

(1) Behavioral bias/ Content production biases: This category describes several bias effects how data and variables are chosen, measured, and presented.
(2) Aggregation bias/Longitudinal data fallacy: Occur during the sampling and analyzing process leading to the fact that estimations which are made for one population may not be transferable to another population.
(3) Ranking bias/Emergent bias: These biases occur when it comes to real user interaction and result from different cultural values, societal knowledge, personal habits etc.

Research shows that humans tend to rely on AI outcomes and have a blind trust in automation (Lee & See, 2004). Further studies also confirm that people are unable to detect algorithmic errors and trust algorithms that are described as accurate but present random results (Rastogi et al., 2022).

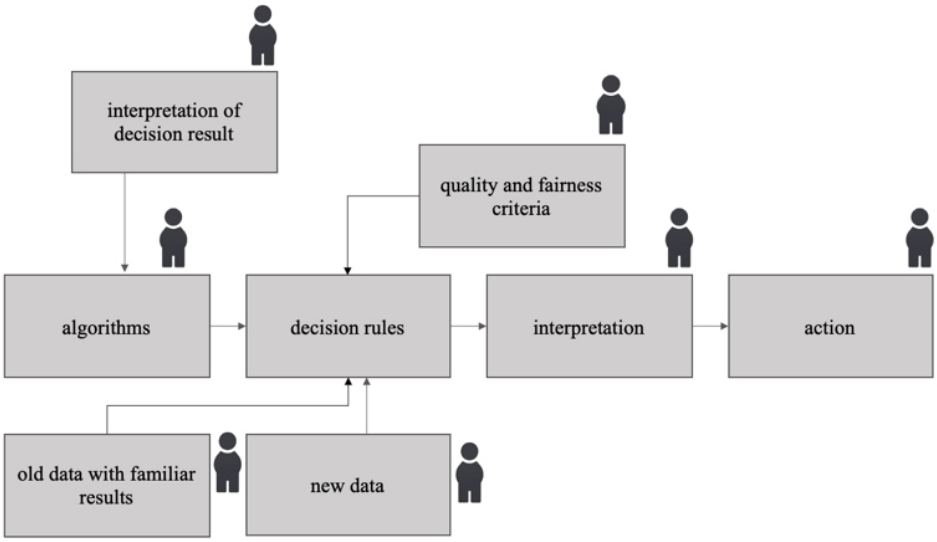

Figure 1: AI and human interaction cycle according to (Zweig & Wilhelm Heyne Verlag Muñchen, n.d.).

In Figure 1 we extend an established interaction cycle to show the function and the interaction with the human of an AI and where in this process cognitive biases can occur (Zweig & Wilhelm Heyne Verlag Muñchen, n.d.).

Every step with a human shows the requirement of human involvement in the usage of AI steps, such as controlling and result interpretation. As human behavior is biased especially in these steps cognitive biases freuquently can occur, and can also result in a biased AI algorithm. Thus, these steps should be taken carefully into consideration to avoid biases in the algorithms when designing AI algorithms (Mehrabi et al., 2019).

METHODOLOGY

To contribute to our research goal, we applied a two-fold research design. First, we combined the literature streams of AI and cognitive biases by

conducting a systematic literature review to show how an AI algorithm works and how cognitive biases influence decisional outcome.

In order to assess the current research perspectives we applied Scopus as primary identifier of relevant literature. Utilizing the following query:

Human
AND decision
AND AI OR artificial AND intelligence
AND bias OR cognitive AND biases

We were able to asses 66 research articles in total. Including articles published between 1990 and 2023, a wide variety of research streams and ages were considered. However, we limited the results to only include journal articles, putting an emphasis on peer-reviewed research. After careful consideration and reading of all identified research articles, a total of 25 research articles are relevant for our research objective. These articles are mainly based on structured literature reviews with additional expert interviews. Seven articles use quantitative experiments for hypotheses testing.

Second, we discussed the impact of biases on artificial intelligence in the field of complex production environments within semi-structured interviews with six experts. The interviews were conducted in German and took between 30–60 minutes. All interviews were transcribed. Based on the expert interviews we iteratively extended the list of relevant cognitive biases and their impact on AI algorithms.

BIASED AI DECISIONS IN COMPLEX PRODUCTION ENVIRONMENTS

Based on our structured literature review as well as on the expert interviews we provide first insights on biased AI decisions in complex production environments. Biases can occur in different steps within the process of working with an AI technology. Therefore, we structured our findings according the classification of cognitive biases, presented in Figure 2.

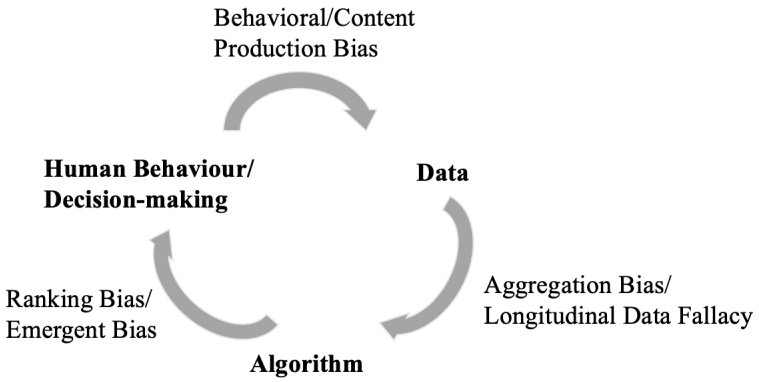

Figure 2: Occurrence of cognitive biases in AI and user interaction according to (Mehrabi et al., 2019).

Behavioral/Content Production Biases

They mainly occur when an AI-Algorithm is created from a limited data set. We identified several potential biases which are active in this phase and can lead to distorted AI based decision-making in production decisions. The measurement bias can arises based on the choice and the definition of the relevant data. This is especially relevant for the evaluation of suppliers. If the products of one supplier are more frequently controlled than the products of another supplier, there could also arise more absolute errors. Therefore, one could not conclude that this supplier delivers worse quality than the other. It is necessary that the same control cycle was applied beforehand.

Aggregation Bias/Longitudinal Data Fallacy

Another frequently active cognitive bias is the so-called representation bias. This effect occurs when the initial population lacks diversity and becomes relevant in AI based decisions, for example in the supplier evaluation and material predictions. It must be assured that there are data from the various types of suppliers ranging from the strategic suppliers as well as from one-time vendors. This is also necessary to avoid a closely related bias type, the so-called aggregation bias which arises when false conclusions about subgroups are made and aggregated to the entire population. For example, based on the production output of one machine type a forecast for the whole production is made. The longitudinal data fallacy arises when data are analyzed for a short period of time and the temporal behavior and development over time is neglected. Especially in the forecast and planning for example of inventory quantities this is crucial.

Ranking Bias/Emergent Bias

A widespread bias effect in this category is the effect of the so-called presentation bias. Based on how information is presented the behavior of the end user is influenced. In the field of the production environment this is an important bias effect which occurs in working with production planning dashboards. Nunc et al. showed in their research about the occurrence of the LTS that active presentation bias influenced decision-making of production planners concerning the update of planned throughput times (Bendul & Knollman, 2016). Another bias effect which is closely connected with the presentation bias effect is the ranking bias effect. This leads to the impression that the ranking or popularity of certain information determines how important there are. This often also occurs in the use of production planning dashboards. One interview partner reported about one case where he assumed that delivery reliability is a more important key performance indicator than the quality measurement, just because there were various figures which reported about the current due date reliability and only one about quality issues.

Based on the various cognitive biases which can distorted AI based decisions we extended the framework of (Lee & See, 2004). We categorized the AI technology according to their damage potentials and their vulnerability for the occurrence of cognitive bias effects. We mapped the first identified examples in these categories and proposed recommendations how to handle

the AI based on the specific category. This could serve as a basis for further research of cognitive biases in the interaction of humans and AI within production environments.

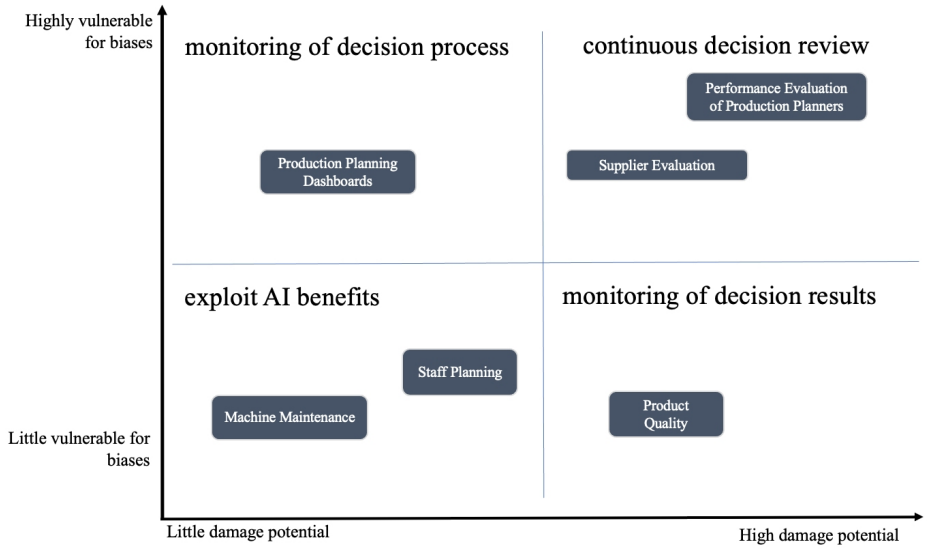

Figure 3: Production control AI bias compass: framework showcasing the bias and damage potentials for AI application in PPC.

CONCLUSION

In this article we showcased first insights into the influence of cognitive biases on AI based decision-making in complex production environments.

Nevertheless, the examples in this article do not claim to be a complete list of all relevant cognitive bias effects. One should consider that our findings are mainly based on discussions with experts coming from the industry, and therefore are also influenced by their respective personal experiences. Moreover, it is important to understand that the classification of biases is not as concrete in practice as described in theory. Some of the effects overlap, and often also occur in several different situations. Therefore, some of our observations might be discarded while others might be missing. Further interviews with experts from various industries could be a promising approach to acquire a broader insight in different AI technologies. To test the influence of the identified biases behavioral experiments could prove a viable avenue for future research.

REFERENCES

Antons, O., & Arlinghaus, J. C. (2022). Data-driven and autonomous manufacturing control in cyber-physical production systems. *Computers in Industry*, *141*, 103711. Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal*, *16*(1), *55*–78.

Arlinghaus, Julia, and Oliver Antons (2022). Management for Digitalization and Industry 4.0. *Handbook Industry 4.0: Law, Technology, Society*. Berlin, Heidelberg: Springer Berlin Heidelberg, 927–948.

Büttner, K., Antons, O., & Arlinghaus, J. C. (2022). Applied Machine Learning for Production Planning and Control: Overview and Potentials. *IFAC-PapersOnLine*, *55*(10), 2629–2634.

Bendul, J. C., & Knollman, M. (2016). The human factor in production planning and control: Considering human needs in computer aided decision-support systems. *International Journal of Manufacturing Technology and Management*, *30*(5), 346–368.

Burggräf, P., Wagner, J., & Koke, B. (2018). Artificial intelligence in production management: A review of the current state of affairs and research trends in academia. *2018 International Conference on Information Management and Processing (ICIMP)*, *2018-January*, 82–88.

Chui L., Kamalnath V., McCarthy B., An Executive's... - Google Scholar. (n.d.). Retrieved January 10, 2024, from https://scholar.google.com/scholar?hl=de&as_sdt=0%2C5&q=Chui+L. %2C+Kamalnath+V. %2C+McCarthy+B. %2C+An+Executive%E2%80%99s+Guide+to+AI&btnG=

Chui M., Intezari A., Taskin N., Pauleen, D. (2021). Cognitive biases in developing biased Artificial Intelligence recruitment system. Proceedings of the 54th Hawai International Conference on Systems Science.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.

Gartner Top 10 Strategic Technology Trends For 2018. (n.d.). Retrieved January 10, 2024, from https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Experiments in Environmental Economics*, *1*, 143–172.

Kamar, E. (2016). *Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence*. https://www.microsoft.com/en-us/research/publication/directions-hybrid-intelligence-complementing-ai-systems-human-intelligence/

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Https://Doi. Org/10.1518/Hfes.46.1.50_30392*, *46*(1), 50–80.

Management and Applications of Complex Systems. (n.d.). Retrieved January 10, 2024, from https://www.witpress.com/books/978-1-78466-367-4.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6).

Microsoft's Bill Gates insists AI is a threat - BBC News. (n.d.). Retrieved January 10, 2024, from https://www.bbc.com/news/31047780.

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 83.

Real-life Examples of Discriminating Artificial Intelligence | by Terence Shin, MSc, MBA | Towards Data Science. (n.d.). Retrieved January 10, 2024, from https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–726.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131.
Windt, K., Philipp, T., & Böse, F. (2008). Complexity cube for the characterization of complex production systems. *International Journal of Computer Integrated Manufacturing*, *21*(2), 195–200.
Zweig K. (2019). Ein Algorithmus hat kein Taktgefühl, Heyne Verlag, München.