Towards an approach for analyzing potentials for development and deployment of human-centered artificial intelligence solutions

Florian BÜLOW[1], Marie-Christin BARTON[2], Simon FAHLE[1], Felix THEWES[3], Bernd KUHLENKÖTTER[1], Jens POEPPELBUSS[2], Thomas HERRMANN[3]

[1] Lehrstuhl für Produktionssysteme, Ruhr-Universität Bochum
Universitätsstraße 150, D-44801 Bochum
[2] Lehrstuhl für Industrial Sales and Service Engineering, Ruhr-Universität Bochum
Universitätsstraße 150, D-44801 Bochum
[3] Lehrstuhl für Informations- und Technikmanagement, Ruhr-Universität Bochum
Universitätsstraße 150, D-44801 Bochum

Abstract. Artificial intelligence (AI) and machine learning are increasingly used in workplaces and human life in general. However, voices from industrial practice as well as academia point to issues with purely technology-oriented approaches to the development and implementation of AI applications. These can include low user acceptance and negative impact on job identification. In response, approaches for a human-centered design of work with AI come into focus. This paper contributes an analytical tool for identifying potentials as a first step in a process model for the development and deployment of AI in a human-centered way. The tool was empirically validated in various real-world settings. The implementation was carried out by the interdisciplinary team of the HUMAINE project consisting of social scientists, engineers, psychologists, information systems researchers and computer scientists. The paper presents the scope, the methods used as well as lessons learned from multiple executions of this concept.

Keywords: human-centered AI, artificial intelligence, AI implementation, sociotechnical systems, analyses of potentials, workflow analysis

1. Introduction

For some years now, artificial intelligence (AI) and especially machine learning (ML) have increasingly found practical application in everyday life and various industries (Littman et al. 2021; Shneiderman 2021; Fahle et al. 2020; Xu 2019). However, voices from industrial practice as well as science recently point to the fact that solely technology-oriented approach to the development and implementation of AI applications can lead to various problems (Xu 2019). These include a negative impact on employees' job identification (Mirbabaie et al. 2021), as well as low user acceptance due to a lack of explainability and accountability (Benbya et al. 2020). In response, approaches to human-centered AI (HCAI) development, implementation and application emerge (Shneiderman 2021; Wilson & Daugherty 2018).

To date, there is no single universally accepted definition of what HCAI is and what aspects it encompasses. However, according to Wilkens et al. (2021b), there are at least five different basic understandings: A deficit-oriented, a data reliability-oriented, a protection-oriented, a potential-oriented and a political-oriented. These different understandings result from the transdisciplinary character of the research field. With disciplines as diverse as psychology, computer science, information systems, engineering, social sciences, and work science looking at the subject, distinct aspects are brought into focus (Wilkens et al. 2021a). Complexity gets even higher when dealing with real-world systems and all their different technical and non-technical components, subsystems as well as various stakeholders from different organizational levels and their interdependencies (Shneiderman 2020). To enable companies and other organizations to anticipate the challenges and opportunities arising from this complex environment already at the very beginning of HCAI system development and implementation projects, suitable methodical support is needed. This paper presents an analytical tool for analyzing the as-is situation, to-be situation, and related potentials of HCAI implementation on distinct levels of an organization.

2. Research Background

There is already a wide range of guidance available for the implementation of AI in organizations. Kaymakci et al. (2021) distinguish between two different groups of approaches: Generic conceptual models of AI systems which include features, functions, and components, as well as blueprints of processes for the implementation, development, and operation of such systems. Despite the multitude of frameworks and processes found in the literature, there is still a lack of approaches that address AI development and implementation in a holistic way. Moreover, most of them relate to a specific industry or application, such as industrial manufacturing (Kaymakci et al. 2021; Pokorni et al. 2021).

For the sake of completeness, it should be mentioned that other tools exist in addition to these groups of approaches identified by Kaymakci et al. (2021). First, there are AI readiness checks in the form of web-based questionnaires or workshops. In some cases, these checks even address different dimensions such as data and technological aspects, human resources, business value, and governance. Then, there are plenty of recommended actions in the form of step-by-step descriptions published by consulting companies and technology blogs. However, these mostly non-scientific approaches are superficial and do not provide methodological support within each step. Finally, there are also less formalized recommendations for AI implementation, which mostly refer to specific aspects. As an example, Davenport (2016) describes distinct ways of introducing AI into organizations, but refers solely to the selection of suitable pilot projects and technology as well as corresponding make-or-buy decisions.

Concluding, it can be stated that the approaches available so far are not suited to cope with the high complexity of HCAI systems (Makarius et al. 2020) due to their limitations regarding their scope, especially when it comes to consideration of individuals and organizational aspects. Thus, for now, a holistic analysis at the very beginning of HCAI system implementation into organizations requires the combination of different methods and frameworks.

3. Scope of the Analytical Tool

The different perspectives identified by Wilkens et al. (2021b) resulting from the transdisciplinary character of HCAI systems indicate which aspects a holistic analysis approach must cover. The deficit-oriented understanding focuses on compensating individual human weaknesses with AI and thus on understanding humans in the context of work. The data-reliability understanding emphasizes deficits in the development of the technology and associated acceptance problems. The political-oriented understanding refers to the distribution of power between humans and AI and thus to the use of the technology, not to its development. The first two, i.e., the protection- and potential-oriented understandings, relate to the physical and mental integrity of human beings and job design, thus addressing human-machine interaction and organizational aspects and are common perspectives taken in work psychology, work science, and engineering. The other understandings appear to be ignored frequently.

It becomes clear, that HCAI systems encompass more aspects than technology and human-machine interaction. According to Ehsan & Riedl (2020), it is necessary to adopt a sociotechnical approach that combines both technical and social elements when dealing with HCAI systems. This includes both the individual background of the users in terms of profession and education, but also the social context of the users in organizational settings with interactions between different users and stakeholders. Makarius et al. (2020) also emphasize the importance of considering social aspects, especially regarding the integration of AI and users within an organization. The corresponding sociotechnical system consists not only of technology and people, but also of the organization surrounding them as well as its structures and processes. We adopt this view and consider HCAI systems as sociotechnical systems with reference to the humans, technology and organization (HTO) concept (Berglund et al. 2020).

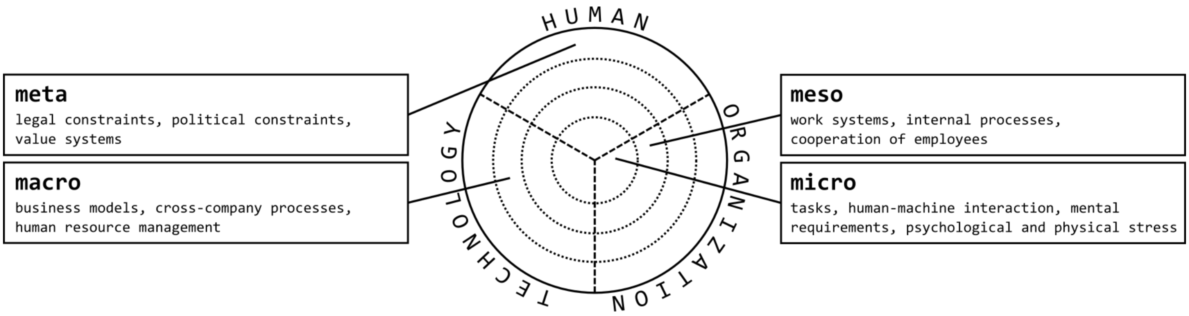

***Figure 1**: Scope of analysis comprising the dimensions of human, technology and organization, subdivided into four levels.*

To structure the analysis, we also consider the dimensions of human, technology, and organization at distinct levels. The micro level of organizations describes individual workplaces. This includes the specific job tasks with their associated mental requirements, as well as the psychological and physical stresses and the human-machine interaction. At the meso level, entire work systems, the processes between different organizational units, and how people collaborate internally are considered. The macro level represents employment arrangements, human resource management, cross-company processes and business models. By considering business models in our analysis, we ensure that not only the potential impact of AI implementation on employees and other stakeholders is considered, but also the competitiveness of the organization as a whole. The organization-related levels are supplemented by a meta level, which describes legal and political constraints as well as overarching value systems (Adolph et al. 2020). In the following, we outline which methods can be used to cover this scope of analysis (see figure 1).

4. Methods used for the Analytical Tool

The analysis starts with the acquisition of a mutual understanding of the relevant business or project. A tool widely used in practice for the analysis, development, and visualization of business models is the Business Model Canvas (BMC). With this strategic management framework, the value proposition, infrastructure, customer segments, financial aspects and thus the underlying business logic of an organization can be described on the basis of various interrelated building blocks (Osterwalder & Pigneur 2010). Thus, the main focus is on the organizational dimension at the macro and meso levels. The BMC can be supplemented by the Value Proposition Canvas (VPC). On the one hand, the VPC helps to obtain a detailed description of the customer segments in terms of their jobs, needs, wishes and problems as well as of the value proposition and its fit with the customer profile (Osterwalder et al. 2014). Besides its application for customer analysis in the context of developing new business models, the method can also be used to create profiles of internal and external users as well as other stakeholders. This is of particular interest on the micro-level in terms of user acceptance towards the AI solutions to be developed or implemented. To identify the stakeholders involved as well as their interrelationships and their influence on the given business model, stakeholder maps can also be used. In some cases, it is meaningful to add the AI Project Canvas (Zawadzki 2019). Designed to assist in pitching AI projects in organizations, this adaptation of the BMC is a suitable tool to further develop existing and described business models with regard to the implementation of AI systems and related implications. The resulting target/actual comparison indicates the need for action to exploit potentials in the organizational dimension.

After a basic understanding of the implementation process at hand has been established by looking at the business model and clarifying the roles of various stakeholders, we suggest using two more technically oriented methods. The methods used are value stream mapping 4.0 and the Cross-Industry Standard Process for Data Mining (CRISP-DM). Value stream mapping 4.0 is an extension of the classic method of value stream mapping. The method can be used to record process data, inventories, control information and the calculation of throughput times. In the course of the emergence of Industry 4.0, it has been extended by digital media and data flows (Mosch & Prumbohm 2018). We suggest to use this method in a lightweight version that only focusses on media discontinuities, process relationships and interfaces within the system, but not on the determination of throughput times and optimization of process planning and control. The CRISP-DM cycle is a process consisting of six key phases within a data science implementation process (Rohanizadeha & Moghadama 2009). We use the method to raise awareness of emerging problems or existing challenges through a moderated discussion about the central issues of the individual phases. Involving all process participants in the implementation of these two tools makes it possible to obtain a holistic and human-centered analysis of the system as well as its technical feasibilities and implementation. Both methods address the macro and meso levels of the technical dimension, with intersections to the other dimensions.

Finally, we use the sociotechnical workflow analysis to analyze current and target states of work processes. This method combines the sociotechnical walkthrough (Herrmann et al. 2007) with a heuristic-based evaluation of sociotechnical systems (Herrmann et al. 2021). We use it to investigate the potentials for improvement and problems occurring during the interaction of humans, technology, and organizational entities as well as intersecting workflows. For the human-centered AI integration, changes in the human-technology interaction and task shifts as well as information flows across organizational units are of particular importance. The sociotechnical workflow analysis also examines the interfaces between the micro and meso levels and their effects on the orientation of the macro level. As a process-oriented method, sociotechnical workflow analysis can be combined with value stream mapping. This combines the data-oriented technical view of value stream mapping with the task-oriented workflow analysis to achieve a holistic perspective.

5. Lessons Learned and Outlook

We have already applied the analytical tool with its methods to various real-world organizations and projects, including a new development in the field of medical imaging diagnostics, the adaptation of an existing system in the field of automotive insurance claims processing, and the development of new AI-based business models in a mechanical engineering company. We applied the approach in guided workshops in which, on the part of the respective organizations, primarily specialists from AI development and data science as well as members of mid-level management participated.

We found that adjustments to the selection and use of methods are necessary depending on the industry, organizational structure, value chain architecture, and project type and stage. For example, if we analyze a development project that is carried out outside the context of an existing organization, e.g. in the run-up to a start-up, the target/actual comparison is omitted and only one BMC is used. We further realized that the quality of the results depends significantly on the industry and professional knowledge of the moderator. Further execution with other organizations will hopefully help us estimating the extent to which this influence can be mitigated, if the analysis can be applied equally well to different types of projects and if there is room for improvement by adapting the methods set. For example, in addition to the described methods, it is planned to integrate psychological questionnaires on job identity as well as a tool for analyzing role development to further explore the human dimension as well as interrelations between the human dimension and the organizational or technological dimension. As for now, the primary benefit of the approach is the identification of interface problems and the establishment of a mutual understanding of the project among various stakeholders. However, in the next development step, it will be important to be able to draw conclusions from these findings and to provide a transition to detailed project planning and implementation.

6. References

Adolph L, Kirchhoff B, Geilen J-H (2020) Sicherheit und Gesundheit in der digitalen Arbeitswelt. In: Maier GW, Engels G, Steffen E (Eds.): Handbuch Gestaltung digitaler und vernetzter Arbeitswelten. Berlin, Heidelberg: Springer, 21–34.

Benbya H, Davenport TH, Pachidi S (2020) Artificial Intelligence in Organizations: Current State and Future Opportunities. SSRN Electronic Journal 19 (4).

Berglund M, Karltun A, Eklund J, Karltun J (2020) HTO - A Concept of Humans, Technology and Organisation in Interaction. HELIX Working Papers. HELIX Competence Centre, Linköping University. Accessed Dec 1, 2021. https://www.diva-portal.org/smash/get/diva2:1511582/FULLTEXT01.pdf.

Davenport TH (2016) 7 Ways to Introduce AI into Your Organization. Accessed Nov 23, 2021. https://hbr.org/2016/10/7-ways-to-introduce-ai-into-your-organization.

Ehsan U, Riedl MO (2020) Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In: Stephanidis C, Kurosu M, Degen H et al. (Eds.): HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence. Cham: Springer International Publishing, 449–466.

Fahle S, Prinz C, Kuhlenkötter B (2020) Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. Procedia CIRP 93, 413–418.

Herrmann T, Jahnke I, Nolte A (2021) A problem-based approach to the advancement of heuristics for socio-technical evaluation. Behaviour & Information Technology, 1–23.

Herrmann T, Loser K-U, Jahnke I (2007) Sociotechnical walkthrough: a means for knowledge integration. The Learning Organization 14 (5), 450–464.

Kaymakci C, Wenninger S, Sauer A (2021) A Holistic Framework for AI Systems in Industrial Applications. In: Ahlemann F, Schütte R, Stieglitz S (Eds.): Innovation Through Information Systems. Cham: Springer International Publishing, 78–93.

Littman ML, Ajunwa I, Berger G, Boutilier C, Currie M, Doshi-Velez F, Hadfield G, Horowitz MC, Isbell C, Kitano H, Levy K, Lyons T, Mitchell M, Shah J, Sloman S, Vallor S, Walsh T (2021) Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Stanford. Accessed Nov 16, 2021. http://ai100.stanford.edu/2021-report.

Makarius EE, Mukherjee D, Fox JD, Fox AK (2020) Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. Journal of Business Research 120, 262–273.

Mirbabaie M, Brünker F, Möllmann NR, Stieglitz S (2021) The rise of artificial intelligence – understanding the AI identity threat at the workplace. Electronic Markets 2021.

Mosch C, Prumbohm F (2018) Leitfaden Industrie 4.0 trifft Lean. Wertschöpfung ganzheitlich steigern. Accessed Nov 10, 2021. https://industrie40.vdma.org/documents/4214230/26095707/Leitfaden_I40_Lean_1524489604061.pdf.

Osterwalder A, Pigneur Y, Bernarda G, Smith A (2014) Value Proposition Design. How to Create Products and Services Customers Want. Hoboken: John Wiley & Sons.

Osterwalder A, Pigneur Y (2010) Business Model Generation. A Handbook for Visionaries, Game Changers, and Challengers. Hoboken: John Wiley & Sons.

Pokorni B, Volz F, Zwerina J, Hämmerle M (2021) Development of a Holistic Method to Implement Artificial Intelligence in Manufacturing Areas. In: Ahram T (Ed.): Advances in Artificial Intelligence, Software and Systems Engineering. Cham: Springer International Publishing, 3–8.

Rohanizadeha SS, Moghadama MB (2009) A Proposed Data Mining Methodology and its Application to Industrial Procedures. Journal of Industrial Engineering 4, 37–50.

Shneiderman B (2021) Human-Centered AI. Issues in Science and Technology (37), 56–61.

Shneiderman B (2020) Bridging the Gap Between Ethics and Practice. ACM Transactions on Interactive Intelligent Systems 10 (4), 1–31.

Wilkens U, Langholf V, Ontrup G, Kluge A (2021a) Towards a maturity model of human-centered AI – A reference for AI implementation at the workplace. In: Sihn W, Schlund S (Eds.): Competence development and learning assistance systems for the data-driven future. Berlin: GITO, 179–198.

Wilkens U, Reyes CC, Treude T, Kluge A (2021b) Understandings and perspectives of human-centered AI – a transdisciplinary literature review. In: Gesellschaft für Arbeitswissenschaft e.V. (Ed.): Arbeit HUMAINE gestalten, Frühjahrskongress 2021, Bochum. GfA-Press.

Wilson HJ, Daugherty PR (2018) Collaborative Intelligence: Humans and AI Are Joining Forces. Harvard Business Review 2018 (4), 114–123.

Xu W (2019) Toward human-centered AI. Interactions 26 (4), 42–46.

Zawadzki J (2019) Introducing the AI Project Canvas. Use the adapted Business Model Canvas to pitch your next AI project. Accessed Dec 5, 2021. https://towardsdatascience.com/introducing-the-ai-project-canvas-e88e29eb7024.

Acknowledgements. This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the “Future of Value Creation. Research on Production Services and Work” (funding number 02L19C200) program and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

Gesellschaft für
Arbeitswissenschaft e.V.

Technologie und Bildung in hybriden Arbeitswelten

68. Kongress der
Gesellschaft für Arbeitswissenschaft

Otto-von-Guericke-Universität Magdeburg

Fraunhofer-Institut für Fabrikbetrieb und
Fabrikautomatisierung IFF, Magdeburg

02. – 04. März 2022

GfA-Press

Bericht zum 68. Arbeitswissenschaftlichen Kongress vom 02. – 04. März 2022

Otto-von Guericke-Universität Magdeburg;
Fraunhofer-Institut für Fabrikbetrieb und -automatisierung IFF, Magdeburg

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Sankt Augustin: GfA-Press, 2022
ISBN 978-3-936804-31-7

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Schriftleitung: Prof. Dr. Rolf Ellegast
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Geschäftsstelle der GfA
Simone John, Tel.: +49 (0)30 1300-13003
Alte Heerstraße 111, D-53757 Sankt Augustin
info@gesellschaft-fuer-arbeitswissenschaft.de · www.gesellschaft-fuer-arbeitswissenschaft.de

Screen design und Umsetzung
© 2022 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de