Towards a maturity model of human-centered AI – A reference for AI implementation at the workplace

Uta Wilkens, Valentin Langholf, Greta Ontrup, Annette Kluge

1. Introduction

Currently there is a third wave in research on artificial intelligence (AI) (Launchbury 2017). In parallel there is a high interest in AI-implementation in organizations (Schuler et al. 2019) and an increasing number of examples where AI is implemented in the workplace (McKinsey 2020; Baruffaldi et al. 2020). Going into more detail it becomes obvious that AI in practice can rather be traced-back to research outcomes from the second wave that emphasized the complementary expertise of AI in correspondence to human intelligence (Brynjolfsson/McAfee 2017; Wilkens 2020) while the third wave in AI research is aiming at an almost perfect copy of human intelligence (Deng 2018). This is an important distinction for workplace analysis and research dedicated to AI implementation.

This paper addresses the implementation of AI at the workplace while suggesting a maturity model of human-centered AI that elaborates on already existing maturity models. The outline refers to typical use cases for current implementation of AI and thus goes beyond the industrial sector. The core emphasis lies on AI-based functions such as enhancing precision, supporting quality control or decision making, protecting security etc. which matter for a variety of work settings in high-tech environments across certain industries. This variety of use cases also increases the number of disciplines involved in the implementation process of AI and leads to different and co-existing interpretations of what human-centricity exactly means and implies for job design. This is why the blueprint includes and combines certain dimensions and criteria indicating how to operationalize the human-centricity of AI and explores a fan structure of the blueprint that allows to set the focus related to context-specific demands during the implementation journey.

In the following paragraphs we will first give attention to typical outlines of maturity models in the field of digitalization. Moreover, we elaborate on a distinct understanding of the human-centricity of AI in order to specify criteria supposed to operationalize the maturity of a human-centered integration of AI in the workplace. The most challenging part is to not just list criteria which are considered as

DOI: https://doi.org/10.30844/wgab_2021_11

relevant in principle but to give evidence to the interrelatedness of these criterin correspondence to the context of implementation. In order to cope with this challenge we refer to qualitative case descriptions from different work settings.

The fan structure of our approach allows to refer to the sociomateriality of technology (Orlikowski 2007; Orlikowski/Scott 2008) and to avoid a techno-centric perspective.

2. Components of a maturity model of human-centered AI

2.1 The way maturity models work

Maturity models in general are assumed to be a helpful tool for describing the state, potentials, and demands within a functional domain (Wendler 2012). Organizations might draw on maturity models to evaluate their status quo and to encourage and monitor a step-wise further development within an implementation process (Alsheibani et al. 2018; Leineweber et al. 2018). In this regard maturity models can also help to leverage capabilities in a specific domain (De Bruin et al. 2005) and to enhance their strategic potential (Alsheibani et al. 2018). With respect to AI-based human-computer-interaction in job design we can elaborate on and further combine two different types of maturity models. The first type is especially helpful to give attention to a comprehensive view in the implementation process. This can be exemplified with the maturity model from Klötzer/Pflaum (2017; see Figure 1).

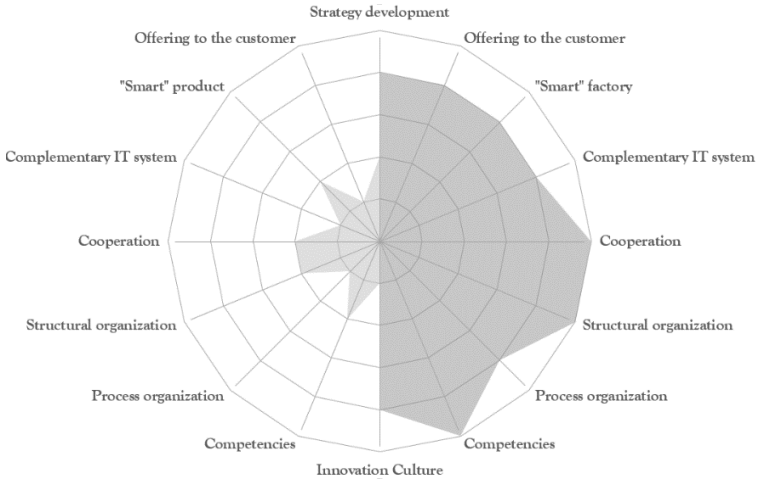

Figure 1: Maturity in the implementation process; own illustration based on Klötzer/Pflaum 2017

The model is not specified for AI but for the digitalization journey of an organization.

The advantage is its comprehensive view of technological, organizational and workforce related components. The shortcoming is that interdependencies between these dimensions are not taken into consideration and that the interrelatedness between technological and human action is not in the center of the approach.

The second type we can elaborate on is represented by the "AI management framework" from Lichtenthaler (2020) as it reflects on the integration of human and artificial intelligence. Low maturity is characterized by experimentations with selected AI technologies in organizations ("initial intent"). High maturity is testified when there is a shared management of human and artificial intelligence ("Intuitive Ingenuity") that involves leveraging pooled, sequential and reciprocal interdependencies between human and AI ("integrated intelligence"). The core argument for the implementation process is that the interrelatedness of AI with the social system is an expression of higher maturity for gaining competitive advantages. The maturity model helps to identify and exploit so far unrealized opportunities while making use of an integrated intelligence structure in organizations (Lichtenthaler 2020). A shortcoming of this model is the missing integration of contextual factors and the underlying assumption that there is one best way to interrelate artificial and human intelligence even though there are co-existing ideas of what human-centricity of AI means (Wilkens et al. 2021). Moreover, the maturity of the technology itself tends to be taken for granted even though there are considerable challenges the implementation process has to cope with.

Level	Name	Description	Impact
+	Intuitive Ingenuity	Shared management of human intelligence and AI; self-awareness systems with some consciousness, emotional intelligence, and ingenuity (only in the future)	Very High
5	Integrated Intelligence	Renewal and recombination of human intelligence and AI; leveraging pooled, sequential, and reciprocal interdependencies for completely novel solutions	
4	Interdependent Innovation	Emphasis on AI for innovation beyond efficiency; sometimes pooled and sequential interdependencies of human intelligence and AI; often corporate orchestration for synergies	
3	Interactive Implementation	Exploitation of multiple AI solutions; sometimes pooled interdependencies of human intelligence and AI; often coordination of activities in multiple organizational units	
2	Independent Initiative	Ongoing AI initiatives; typical emphasis on advanced automation and enhancing efficiency of established processes; often started in selected organizational units	
1	Initial Intent	Initial steps of experimenting with selected AI technologies; exploration of feasibility and viability; limited implementation in uncertain context	

Figure 2: Maturity as integration of human and artificial intelligence; own illustration based on Lichtenthaler 2020

While both introduced models give inspiration maturity models emphasizing the pure technological readiness of AI applications in the workplace (e.g. Aronsson et

al. 2020) represent a less helpful direction for the aim of this paper as they have low potential to face human-centricity of technology in an adequate manner.

To sum up, so far there is no maturity model reflecting a deeper understanding of what human-centricity of AI means and that at the same time allows to reflect on the technological advancement and to come in close with a comprehensive view of technology implementation.

2.2 The (multiple) perspectives of human-centricity of AI

There is a wide range of interpretations across disciplines of what human-centered AI exactly means. In their literature review Wilkens et al. (2021) identify five different perspectives – a deficit-oriented, a data reliability-oriented, a protection-oriented, a potential-oriented and a political-oriented understanding of human-centered AI (see Figure 3).

The **deficit-oriented understanding** of human-centered AI considers AI as beneficial and helpful to compensate individual weaknesses and failure in attention, concentration, physical and mental fatigue. According to this understanding the maturity of human-centricity can be measured with respect to the *failure control function of the technology*. A high maturity level in this regard indicates that technology prevents the human being from time consuming monotonous work and individual fatigue due to these exhausting job characteristics. This concern is often stressed in health care and clinical settings, for example to relieve nurses of constant tracking activities of health parameters (Adnan et al. 2020). An example from production settings is the use of AI for analyzing photographs for quality control reasons of welding seams. Another example in the field of software engineering concerns the design of system interfaces that minimize cognitive load of users (Oviatt 2006).

The **data reliability-oriented understanding** of human-centered AI refers to existing deficits of the AI technology in order to provide a reliable tool in individual decision making. Criteria of human-centered AI referring to this perspective are reliability, validity, fairness and explainability. Due to the complex nature of some machine learning algorithms, i.e. deep learning algorithms, it is important that the scope, type and quality of data used as an input are visible. Further, opacity of algorithmic output and decisions must be avoided (Gal et al. 2020). In this regard, the *explainability* of AI outputs is particularly important for user acceptance (Deng 2018) but most challenging since deep neural networks provide more powerful outputs the more complex their hidden structure is (Rai 2020). Two technical approaches for reaching explainability are "intrinsic interpretability" which aims at developing less complex, self-explanatory models that can be instantly interpreted by humans or "post-hoc interpretability methods" that aim at explaining very complex models by describing the logic through similar less complex models (Bauer et al. 2021). Further, explainability can be aimed at enabling a global understanding of the models (its structures, assumptions, parameters) or local interpretations

(contribution of input features to output; Bauer et al. 2021). The basic approach is to systematically analyze the hidden structures of the neural networks using suitable tools in such a way that, in addition to the actual outputs in terms of decision support, further information are provided to the users that supports and empowers them to understand the decision criteria of the model. Human-centricity is reached when a user can recognize biases or quickly detect causes for certain misjudgments of the model (Meske et al. 2020). The disclosure and explanation of decision criteria enable humans to test for possible violations against commonly agreed upon (and potentially legally anchored) fairness criteria like avoiding discrimination. The disclosure of decision criteria gives the human information about the reliability of the AI-based decision results and may lead to a trustful relationship. In this context, explainability research focusses on how humans respond to different types of explanations (e.g. intuitiveness, usability; Bauer et al. 2021). Explainable AI is thus considered a multidimensional challenge, as it involves not only technical trade-offs between prediction accuracy and transparency of results but also political and societal efforts (Beaudouin et al. 2020).

The **protection-oriented understanding** of human-centered AI focuses on the physical and mental integrity of the human being. Job design and technology development follow the principle of providing work contexts with tasks that are executable, harmless and safe, tolerable and personality enhancing (Hacker/Richter 1980). AI applications may be implemented for the reason of protecting workers from possible risky and harmful working conditions (Giusti et al. 2018). AI applications in mobile robots can prevent humans from working in environments with e.g. high radiation or other forms of contamination. A mobile robot can carry a load that would be too heavy to carry by humans. An AI expert of DHL explains (Port of Rotterdam 2021) that robots can take over hazardous tasks from humans, don't mind monotonous, repetitive work and can carry more load. The Port of Rotterdam also uses AI-based Automated Guided Vehicles which take care of transport of containers to the depot. They are unmanned, fully automated and recognize by themselves when their battery is almost empty. They then drive to the battery exchange station and receive a new battery from a robot. The core issue of the protection-oriented perspective is to free employees from heavy loads, possible harm and intolerable working demands. Human wellbeing is in the center of the optimization.

Taking the burden off human workers is the first step, but protection-orientation does not refer solely to the release from health damages but also from bad work design with risks involved (e.g. due to skill decay, dissatisfaction and frustration, no commitment and fear of social isolation) of the work portions that remain for the human worker. These aspects are represented in the model of SMART work design (Parker/Grote 2020). Human-centered AI can contribute to these aspects by providing motivating work settings that strengthen human agency and are meaningful and challenging for all humans involved.

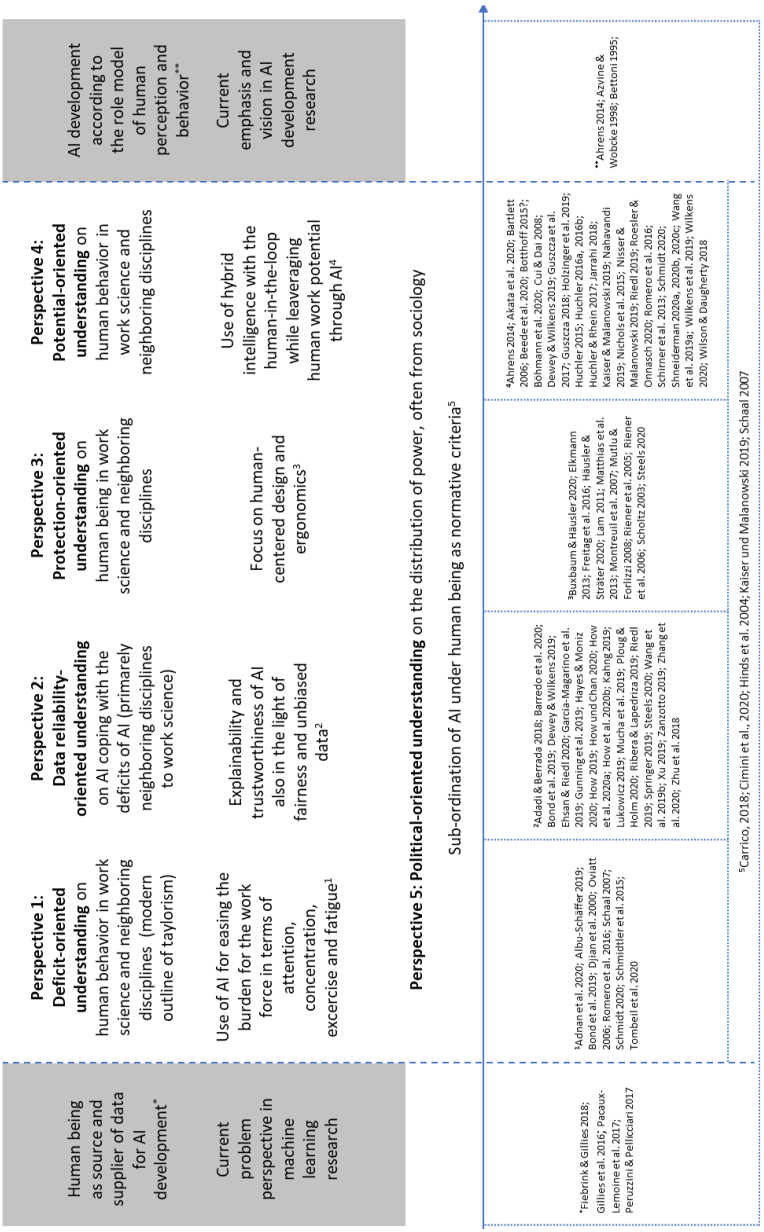

Figure 3: Five perspectives on the meaning of human-centered AI
list of references see: https://seafile.noc.ruhr-uni-bochum.de/f/d636e976042347b5af09/

The **potential-oriented understanding** of human-centered AI gives emphasis to a so far unexploited potential of leveraging individual abilities while developing work systems with hybrid intelligence bringing together individual intelligence with AI in a collaborative manner. There is a strong belief in better outcomes for individual and organizational development as well as task proficiency at the same time. Human-centered AI thus means that technology is utilized in a way that is beneficial for the *competence development* and learning of users (Vladova et al. 2019). Whether competences can be enhanced through AI or even reduced is a matter of socio-technical system design (Wilkens et al. 2019). Another concern within the potential-oriented understanding is *work design based on strengths of humans and technology*. Human-centricity is reached by creating work systems where humans and technology can "complement each other through human-computer cooperation" (Cui/Dai 2008). In that sense, the aspired goal is to develop AI solutions which can also be described as human-AI teaming, as active and interdependent collaborations between humans and AI to achieve a common goal (O'Neill et al. 2020). This includes opportunities for communication between humans and AI, mutual support, shared understanding of the situation or a mutual recognition of intentions (Chen et al. 2018). Another important focus within the potential-oriented understanding is on work systems that benefit from *distributed intelligence* (Fischer 2001; Cobb 1998).

The **political-oriented understanding** of human-centered AI gives emphasis to the distribution of power between AI and those who use AI in the work context. The main concern in this regard is that AI remains under human control. This perspective is applied to research on robot design that ensures human responsibility (Hinds et al. 2004). Other research raises awareness for social aspects, when an increasing number of robots with different roles is used in socio-technical systems in the manufacturing industry (Moniz/Krings 2016). The main concern of this perspective is establishing *regulation for subordinating technology under individual control.*

Based on these perspectives, human-centricity of AI solutions are reflected in various related concerns. These concerns go beyond the consideration of encompassing integrations of human and artificial intelligence, which for example the "AI management framework" would consider the highest level of maturity (Lichtenthaler 2020). Rather, the design and implementation of human-centered AI solutions implies to consider technological (e.g. securing reliability), human (e.g. ensuring protection) and organizational (e.g. reducing deficits and enhancing potential) requirements.

It becomes obvious from the overview that there are many questions how to relate these criteria to each other and that some of them even might contradict to each other especially in the field of how to manage control. This is why we try to gain a deeper understanding from selected use cases which allow to better understand how to exploit the potential of AI in a human-centered manner.

Table 1 gives a brief summary of the specific concerns of each of the five perspectives of human-centered AI.

Perspective of human-centered AI	**Concerns of the perspective**
Deficit-oriented	Failure control function of the technology
Data reliability-oriented	Disclosure of decision criteria by AI System Testing for fairness and reliability by human worker Aiming at a transparent outputs for the user
Protection-oriented	Ensuring that work is executable, harmless and safe Enabling work that is stimulating and personality enhancing
Potential-oriented	Enabling competence development Work design based on strengths of humans and technology Work systems with active and interdependent collaborations between humans and AI (distributed intelligence)
Political-oriented	Regulation for subordinating technology under individual control

Table 1: List of concerns for the five perspectives of human-centered AI

3. Qualitative case descriptions

3.1 Insights from practice - what do we learn from good and bad practices?

In this paragraph we introduce selected cases with reference to the five perspectives of human-centered AI in order to better understand how to indicate maturity and how to relate certain criteria to each other. The case selection follows the principle to address those fields where AI enters the workplace and unfolds a specific function such as enhancing precision, supporting quality control or decision making, facilitating learning or carrying heavy loads. In this regard there are similarities between workplaces from different industries and sectors. This is why the case selection goes beyond manufacturing and also includes the healthcare or training sector. The descriptions are derived from interview studies or daily exchange with practitioners respectively from case descriptions in the literature.

Example 1: AI-based diagnosis in radiology

There is an increasing emphasis on precision and quality control enhancement in the field of medicine, especially radiology (Thrall et al. 2018), where AI is in use for diagnosis and treatment suggestions with feedback-loops between therapy and

diagnosis. There are similar developments in AI-based imaging in steel production but the evaluation of the consequences for job design is more advanced in medicine and thus in the focus of the exemplification. Interview studies with radiologists show that physicians especially appreciate AI as a tool that allows them to get rid of monotonous tasks and to gain better output in decision making processes:

"I believe at some point the attention threshold is simply no longer the same after five hours as it was after the first hour. And I think that's what it's good for. If a machine learning program runs in the background like a safety net and really displays "So, I find this striking, don't you want to look at it again?" Or perhaps during the shifts in the hospital, the radiologists are not on site 24 hours a day, that the clinicians in the emergency department justify their images themselves and make decisions, and if they are young colleagues, then they simply haven't seen so many images yet, and I believe that machine learning can be a good support as a safety measure." (see interview study from Wilkens/Langholf 2021).

"Where I can also imagine it well is, for example, when they do a staging examination of the lungs. There, it's often just a matter of counting the metastases and that's not very exciting for us and more of a hard work. So I can also imagine that the AI will do it in the future. Yes, I think the risks are just, above all, that the algorithms are not so good, because they are only tested on their data set with which the algorithm is developed. It actually has to be fed more continuously in order to be as usable as possible." (see interview study from Wilkens/Langholf 2021).

Radiologists make also clear that they first expect the trustworthiness of AI as a prerequisite before making use of it – this might be especially relevant when there is a high responsibility for human life:

"We know that from many research projects that students sometimes do the annotation because it's inexpensive. But of cause this is problematic. The radiologist should actually do the annotation so that there is a well performing AI in the end. And this is something I often see: When the algorithm is written poorly the product will be poor as well. Then it won't be of real use in the hospital after all." (see interview study from Wilkens/Langholf 2021).

According to this case the failure control function of technology for compensating individual deficits or failure is highly appreciated as an issue of enhancing the individual expertise. There is no fear that the individual status could suffer but quite the opposite that the individual expert status could benefit from better decision making. However, the prerequisite and necessary condition for making use of AI tools is its reliability and the elaboration on trustworthiness. Otherwise there would be no basis for enhancing expertise.

Example 2: Radiographers and speech therapists unlikely working with the AI machine

There are job designs for human-computer-interaction very close to the described workplace of radiologists but however lead to different perceptions and attributions from the employees. As a typical radiographer working at the same place as the radiologists argues:

"The computer takes everything off our hands, it already places everything and I actually only have to say okay. [...] I see [...] that I am becoming more dispensable." (see interview study from Wilkens/Langholf 2021).

In a similar manner there was the idea of AI developers to introduce an AI-based logopedic training system. With the AI training system the patients should be able to practice independently at home. The AI system was able to give individual feedback and correct pronunciation. The goal was to allow patients to continue practicing between sessions with a speech therapist and between face-to-face meetings and thus make faster progress. In that way, the AI solution could be used to compensate for a lack of resources (time, availability) of the speech therapists. Even though one might assume that the advantages for the patients are obvious – as the system can be considered as a supportive training aid - the AI-based speech therapy training system was not in use by the speech therapists. They were in concern that the AI system would replace them one day and this was weighted much higher than the possible benefit for patients.

These examples show that in these cases the AI users in the workplace are afraid of losing expertise and are not involved when considering future perspectives of organizational development including individual job profiles. This fear of losing expertise can also be expected for tasks profiles in industry which are based on vocational training. It becomes obvious from these case descriptions that regulations for subordinating technology under individual control define a necessary condition for otherwise less involved employees. In addition to the trustworthiness of AI this tends to be another prerequisite in order to make use of AI in the workplace.

Example 3: Experts in quality management searching for a new role concept

A petrochemical company introduces an AI system that analyzes photographs made by a drone that show welding seams of the pipes of a chemical plant. Weldings seams need to be checked for possible porous and cracked parts. In the past, an expert took the photos with a camera and analyzed them regarding these porous and cracked parts. The task requires extensive vocational training and certified expertise in the quality control of the pipes welding seams. Previously, the job fulfilled the criteria of human-centered work design in terms of identity, agency and satisfaction. After the implementation of the AI based visual quality inspection, there is the risk assigning those aspects of the former well designed task. If the AI would only ask the human in cases of inconclusiveness the human would be in the role of supervising the AI over a longer period of the day, interrupted by some troubleshooting tasks, in the case that the AI is helpless. This kind of implementation is violating the criteria of identity, agency and satisfaction.

It becomes obvious from the example that the integration of AI needs to be combined with a role development perspective for the employees in order to fulfill

workers' needs with respect to agency and stimulating work that leads to the perception of mastery.

Example 4: AI-based tools for individual competence development and career planning

A well-known and sophisticated tool for AI-based individual competency management and career development is the example of the IBM Watson Career Coach (see Guenole/Feinzig 2018). The general idea is that the „Watson Career Coach", as a trusted AI advisor, is consulted by employees for career advice. The organization thus requires competence- and career-specific data of company/competence profiles as well as job profiles, information on how long a person has been with the company and also in a specific position, comparison with data of other job holders in the same position. The company-specific trained AI based career coach "Watson" learns what moves and interests the employees in the company. Subsequently, the career coach simulates the next career step of each employee. In contrast to human resource personnel or supervisors, Watson is able to integrate big data information – arguably with less bias – and might thus provide employees with more objective and reliable career advice that has more predictive value than intuitive predictions.

The capacity of AI to enhance and support individual potentials becomes obvious from this example. Yet, it is still far from practical application as missing data quality is as severe as in the field of hospitals as described in the first example. This is especially the case because companies' practices in career development were not free from discrimination in the past. Available data do not necessarily lead to fair practices in future (violating the criterion of reliability-oriented AI). This means that the trustworthiness of AI and the disclosure of decision criteria is a key prerequisite also for enhancing potential.

3.2 Blueprint for a maturity model of human-centered AI

Our aim is to stress that due to the five different perspectives of human-centricity, maturity is not achieved by fulfilling each and every one of the criteria but by establishing configurations of these criteria that represent a balanced and context-related approach. The introduced case descriptions underline that it is in principle more than one criterion that matters but that a context related selection and focus tends to be helpful at the same time. A maturity model is a suitable guideline for the implementation process if it allows to monitor the most important firm specific criteria without neglecting complexity.

The first outcome from the case description is that the trustworthiness of AI and the disclosure of decision making criteria are necessary conditions for integrating AI in job designs with human-computer interaction. This became obvious when radiologists explained that they otherwise would not make use of the technology because they have to take the responsibility for the therapy. It became also obvious

from the example of the career coach. If the individual development depends on AI-based support, the individual needs to be convinced that there is no hidden discrimination due to faulty algorithms. Trustworthiness can result from the visibility of the input (scope, type and quality of data) and the explainability of the output/ decisions. It equally matters for fields where AI compensates human deficits and enhances individual competence development.

The second outcome is that the integration of AI creates a vacuum for employees' upward or downward development especially in fields with high expertise and proficiency in a non-academic manner but based on vocational training. This became obvious when radiographers or speech therapists expressed their concern of losing their expert status. This is why there is a need for regulation which takes into consideration the individual involvement in workplace development including the moderation of a role development process. This is a further necessary condition for a human-centered way of integrating AI in the workplace.

The third outcome related to the protection-oriented perspective is that job design should not only include criteria where AI can prevent from physical and mental harm – this defines an overall necessary condition for job design – but also consider criteria how AI can contribute to a personality supportive job design in terms of identity, agency and satisfaction. This became obvious from the third example referring to quality control in welding seems. The individual job profile needs to be coherent and integrative and cannot be reduced to troubleshooting functions. This supposed to describe a sufficient condition.

The fourth outcome is that there are further dimensions of a human-centered AI that can be classified as sufficient conditions for the use of technology. These dimensions are not necessarily interrelated but define alternative ways of using AI for human-computer interaction. One direction was just mentioned in the field of personality development in quality control. Two other directions result from the deficit-oriented and the potential-oriented perspective. There are job profiles, e.g. for physicians in radiology, where the compensation of failure is appreciated with respect to the quality of output and the attribution of individual expertise. As the first example from radiology explored it was the compensation of deficits which was considered as basis for enhancing the expert role of radiologists. There are other job profiles where the enhancement of individual expertise through the use of AI can be considered as beneficial for the proficiency of output and the individual development perspective, e.g in AI-based career development. Table 2 summarizes these criteria and relates them to different maturity levels. The blueprint specifies criteria for all five dimensions that are based on the perspectives of human-centered AI. In addition, all criteria are organized in a logical hierarchical order indicating different levels of maturity.

According to the examples chosen in this outline, there is some evidence that it depends on the context which criteria are more likely in the focus. It might be a

combination of regulation and compensation of deficits together with personality enhancing job design especially in manufacturing but a combination of trustworthiness and potential-enhancement in the field of services.

	Dimension	Criteria (how to reach)
Necessary conditions	Trustworthiness and explainability of AI	T1: Availability of big data T2: Cleanup of data (scope, type and quality of data) T3: Explainability of decisions (disclosure of data structure in decision support) T4: Integration of (implicit) user-domain knowledge in data management
	Regulation (Degree of regulation for subordinating technology under individual control)	R1: AI users informed about future workplace perspectives R2: Users involved in workplace development R3: Role development concept specified R4: Role development concept ratified in labor regulation
	Protecting individuals	P1: AI detects if criteria of tolerable work (executable, harmless, safe) are hazarded P2: AI ensures that criteria of tolerable work can be fulfilled
Sufficient conditions	Personality enhancing	P3: AI enhances personality-supportive job design (identity, agency and satisfaction) P4: AI enhances social job design (stimulating, mastery and agency enhancing)
	Compensating deficits	C1: AI detects failure automatically C2: AI informs the user C3: AI prevents from repeating failure C4: AI initiates feedback-loops on system level
	Enhancing potential	E1: AI provides impulses to individual competence development E2: AI integrates human intelligence in critical processes E3: AI creates a working system with collaborating humans and AI (distributed intelligence).

Table 2. Blueprint: Dimensions and criteria for human-centered AI maturity

To account for this context dependency, the blueprint can be understood as a fan model (Figure 4). The model proposes a dynamic, context-sensitive implementation of human-centered AI. The rotatable outer circle implies that the focal sufficient conditions of human-centered AI can be different depending on the specific context of AI implementation. The dashed lines symbolize the foldable nature of the conditions, which, like a fan, allow to map a context-appropriate configuration of conditions for human-centered AI.

The advantage of the introduced model is that it allows to integrate context factors with respect to the concrete workplace design and avoids complexity where unnecessary for the implementation journey. In this regard our approach reflects on the sociomateriality (Orlikowski/Scott 2008) and goes beyond the models introduced in paragraph 2.1. The blueprint of the fan model indicates that there are certain ways how to refer to the interrelatedness between technology and human

beings and that this is crucial for monitoring the human-centricity. To further illustrate this consideration we point to two specific cases outlined earlier and show how the fan model can be applied to these specific contexts.

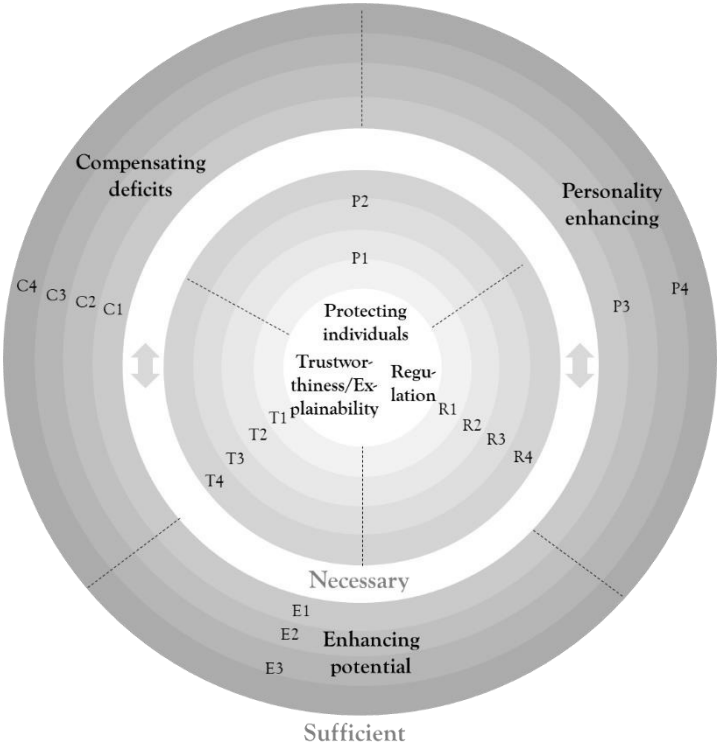

Figure 4: Blueprint: fan model for maturity in human-centered AI implementation

The radiologists (Example 1) had a very specific concept of AI in their work system due to their background and training. The prior role of humans was taken for granted and there was no concern about missing a personality-related job design but trustworthiness of AI was a crucial point for them to make use of the technology. This is why it defines a dominant necessary condition. Another key concern of radiologists was that the technology eliminates human error and relieves them of monotonous work. This concern is related to the deficit-compensating function of AI which is considered as job design making the expert role of radiologists even stronger (see Figure 5, left side).

From Example 2, the speech therapy system, it was not the trustworthiness of the technology which was in concern but the remaining vacuum with respect to the own job profile. This is why the regulation with respect to a personality-enhancing

job profile turned out to be particularly important criteria which defines a necessary condition which the personality-enhancing components themselves define a sufficient condition (see Figure 5, right side).

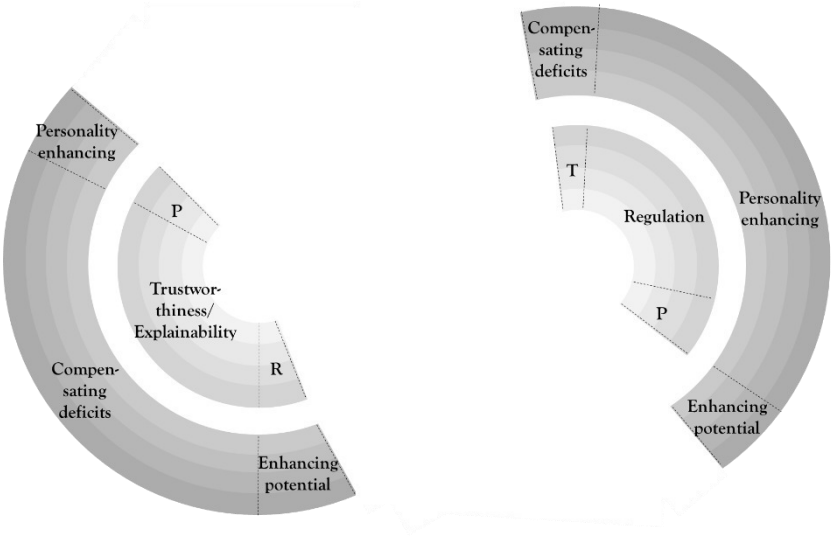

Figure 5: Fan model applied to radiologists (Example 1, left side) and speech therapists (Example 2, right side)

4. Summary, Outlook and Limitations

In this paper we suggested a blueprint for AI implementation in the workplace which includes two aspects. It takes into consideration certain definitions and perspectives of what human-centered AI means and gives evidence to these perspectives in a context-specific manner. This is why the model includes complexity on the one hand side but encourages to reduce this complexity for a context-specific implementation journey. Hence, the blueprint works as a fan model.

The blueprint suggests different maturity levels and how to operationalize them for each of the five dimension included either as a necessary or as a sufficient condition. Whether this operationalization is coherent and also balanced with respect to the distances between the maturity levels needs to be validated in a simulation environment and by the help of a quantitative test design in future research. An important next step could be laboratory studies e.g. with radiologists or students from medical school who have to make decisions by the help of AI-based imaging, how they reflect on the decision support, how they evaluate the explainability and how they estimate their expert role etc. This is also important in order to learn more about the relationship and interrelatedness between the proposed

maturity criteria of successful transformation. Moreover, the validation process can help to identify criteria which might have been neglected so far. But the most interesting and promising part of the empirical exploration is to find out whether there are clusters and core combinations with high relevance for certain fields of AI implementation or combinations of criteria which would contradict each other.

Another most important prerequisite for initiating empirical testing is the question whether the fan model idea has an intuitive plausibility for stakeholders who are involved in the AI implementation process. The floor for this discourse is now open.

References

Adnan, H. S., Matthews, S., Hackl, M.; Das, P. P., Manaswini, M., Gadamsetti, S., Filali, M., Owoyele, B., Santuber, J., & Edelman, J. (2020). Human centered AI design for clinical monitoring and data management. *European Journal of Public Health, 30*(5).

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial Intelligence Adoption: AI-readiness at Firm-Level. In PACIS (p. 37).

Aronsson, J., Lu, P., Strüber, D., & Berger, T. (2020). A Maturity Assessment Framework for Conversational AI Development Platforms. arXiv preprint arXiv:2012.11976.

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, d. & Squicciarini, M. (2020). Identifying and measuring developments in artificial intelligence: Making the impossible possible. *OECD Science, Technology and Industry Working Papers, 2020/05*, Paris: OECD Publishing.

Bauer, K., Hinz, O., van der Aalst, W., & Weinhardt, C. (2021). Expl (ai) n it to me–explainable AI and information systems research. *Business & Informations Systems Engineering 63*(2):79–82. https://doi.org/10.1007/s12599-021-00683-2

Beaudouin, V., Bloch, I., Bounie, D., Clémençon, S., d'Alché-Buc, F., Eagan, J., Maxwell, W., Mozharovskyi, P., & Parekh, J. (2020). Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach. *hal-02506409*

Brynjolfsson, E., & Mcafee, A. N. D. R. E. W. (2017). The business of artificial intelligence. *Harvard Business Review, 7*, 3-11.

Chen, J. Y., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science, 19*(3), 259-282.

Cobb, P. (1998). Learning from distributed theories of intelligence. *Mind, Culture, and Activity*, 5(3), 187–204.

Cui, X., & Dai, R. (2008). A human-centred intelligent system framework: meta-synthetic engineering. *International Journal of Intelligent Information and Database Systems 2* (1):82–105

De Bruin, T., Rosemann, M., Freeze, R., & Kaulkarni, U. (2005). Understanding the Main Phases of Developing a Maturity Assessment Model. In Bunker, D, Campbell, B, & Underwood, J (Chair), *Australasian Conference on Information Systems (ACIS*). Symposium conducted at the meeting of Australasian Chapter of the Association for Information Systems.

Deng, L. (2018). Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [perspectives]. *IEEE Signal Processing Magazine, 35*(1), 180-177.

Fischer, G. (2001). Communities of Interest: Learning through the Interaction of Multiple Knowledge Systems. *24th Annual Information Systems Research Seminar in Scandinavia (IRIS'24) (Ulvik, Norway), Department of Information Science, Bergen, Norway*, 1-14.

Gal, U., Jensen, T. B., & Stein, M. K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization, 30*(2), 100301.

Giusti, A., Steiner, D., & Bertoli, S., & Matt, D. T. (2018). Entwicklung eines flexiblen, inkrementell lernenden Programmiersystems für kollaborative Roboterapplikationen. In: D. T. Matt (Ed.), *KMU 4.0 - Digitale Transformation in kleinen und mittelständischen Unternehmen* (p. 233–248). Berlin: GITO-Verlag.

Guenole, N. & Feinzig, S. (2018). *The business case for AI in HR.* IBM Corporation

Hacker, W. & Richter, P. (1980). *Psychische Fehlbeanspruchung, psychische Ermüdung. Monotonie, Sättigung und Stress.* Bern: Huber.

Hinds, P., Roberts, T., & Jones, H. (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Comp. Interaction 19* (1):151–181.

Launchbury, J. (2017). A DARPA perspective on artificial intelligence. *DARPA talk*, February 15, 2017.

Leineweber, S., Wienbruch, T., & Kuhlenkötter, B. (2018). Konzept zur Unterstützung der Digitalen Transformation von Kleinen und Mittelständischen Unternehmen. In: D. T. Matt (Ed.), *KMU 4.0 - Digitale Transformation in kleinen und mittelständischen Unternehmen* (p. 21–39). Berlin: GITO-Verlag.

Lichtenthaler, U. (2020). Five Maturity Levels of Managing AI: From Isolated Ignorance to Integrated Intelligence. *Journal of Innovation Management, 8*(1).

McKinsey (2020). *The state of AI in 2020*. Available online: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020 (accessed on April, 30, 2021).

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2020). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*. Advance Online Publication.

Moniz, A. B., & Krings, B. J. (2016). Robots working with humans or humans working with robots? Searching for social dimensions in new human-robot interaction in industry. *Societies, 6*(3), 23.

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020) Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors* (Advance online publication).

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies, 28*(9), 1435-1448.

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: challenging the separation of technology, work and organization. *Academy of Management annals, 2*(1), 433-474.

Oviatt, S. (2006). Human-centered design meets cognitive load theory: designing interfaces that help people think. *Proceedings of the 14th ACM international conference on Multimedia*, 871–880.

Parker, S. K., & Grote, G. (2020). Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Applied Psychology: An International Review, 2020, 0 (0), 1–45.

Port of Rotterdam (2021). Der Roboter ist im Anmarsch. URL: https://www.portofrotterdam.com/de/geschaeftsmoeglichkeiten/logistik/ladung/container/50-jahre-container/der-roboter-ist-im-anmarsch, Retrieved 24.05.2021

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137-141.

Schuler, S., Hämmerle, M., & Bauer, W. (2019). Einfluss Künstlicher Intelligenz auf die Arbeitswelten der Zukunft. In: D. Spath & B. Spanner-Ulmer (Eds.), *Digitale Transformation. Gutes Arbeiten und Qualifizierung aktiv gestalten* (p. 255–272). Berlin: GITO-Verlag.

Thrall, J. H., Li, X., Li, Q., Cruz, C., Do, S., Dreyer, K., & Brink, J. (2018). Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology, 15*(3), 504-508.

Vladova, G., Gronau, N., & Rüdian, S. (2019). Wissenstransfer in Bildung und Weiterbildung: Der Beitrag Künstlicher Intelligenz. In: D. Spath & B. Spanner-Ulmer (Eds.), *Digitale Transformation. Gutes Arbeiten und Qualifizierung aktiv gestalten* (p. 89-106). Berlin: Gito-Verlag.

Wendler, R. (2012). The maturity of maturity model research: A systematic mapping study. *Information and software technology, 54*(12), 1317-1339.

Wilkens, U. (2020). Artificial Intelligence in the workplace – A double-edged sword. *International Journal of Information and learning technology, 37*(5), 253-265.

Wilkens, U., Cost Reyes, C., Treude, T.; Kluge, A. (2021). Understandings and perspectives of human-centered AI. A transdisciplinary literature review. GfA, Dortmund (Hrsg.): *Frühjahrskongress 2021, Bochum*, Beitrag B.10.17.

Wilkens, U., & Langholf, V (2021). How to come in close with human-centered AI in the workplace? Insight from a field study analysis in radiology. *83. VHB Jahrestagung, Track: Personal, Düsseldorf*, March 8 - 11, 2022.

Wilkens, U, Lins, D., Prinz, C., & Kuhlenkötter, B. (2019). Lernen und Kompetenzentwicklung in Arbeitssystemen mit künstlicher Intelligenz. In: D. Spath & B. Spanner-Ulmer (Eds.), Digitale Transformation. Gutes Arbeiten und Qualifizierung aktiv gestalten (p. 71–88). Berlin: GITO-Verlag.