Reconfiguring Workers’ Expectations on AI Reliability: The Evolvement of Trust during AI Implementation

*Amelie Schmid, TU Dortmund University & Robert Bosch GmbH,*
amelie.schmid@tu-dortmund.de

*Manuel Wiesche, TU Dortmund University,* manuel.wiesche@tu-dortmund.de

**Introduction and Motivation**

Artificial intelligence (AI) as a new driver in the field of technology has a major impact on the nature of work, e.g. in the way organizational decisions are made or in the field of knowledge management (Wirtz et al., 2018). To unleash the full potential of AI, as a prerequisite a high degree of trust into the emerging technology is mandatory. However, the lack of trust represents one of the biggest challenges for workers take full advantage of AI (Gillath et al., 2021). In particular, this is valid for the so-called “embedded AI”, which is characterized by being completely “invisible” to workers (Glikson & Woolley, 2020).

There are existing numerous influencing factors for building trust. The positive effect of performance related aspects on the trust in IT systems is already well known (Söllner et al., 2016). It has to be pointed out that the reliability of the IT system serves as an important baseline for building trust in technology. It refers to the confidence of the worker, that the technology will operate in a suitable and consistent way (McKnight et al., 2011). Nevertheless, AI is often described as a “black box”, which is basically caused by the fact that it is difficult for workers to completely understand the algorithm with its underlying rules and the logic behind (Adadi & Berrada, 2018; Hoff & Bashir, 2015; Li & Hahn, 2022). As a result, workers feel to be faced with a non-transparent decision making process of AI and thus have to tackle challenges regarding the right interpretation of AI output (Danks & London, 2017). Based on that it is claimed that the current assumptions regarding the impact of reliability on building trust in technology need an extension towards AI-based systems.

In the past, a rather static relationship between the influencing dimensions of trust and workers’ trust building in technology is assumed (Glikson & Woolley, 2020). However, in the context of AI it is already known that the process of building trust in AI needs to be continuously maintained during its development and implementation (Hengstler et al., 2016). So far, there is no clear understanding with respect to the dynamic process of building trust in AI during AI implementation, as well as the implications for workers. Thus, the major goal of the present study is to investigate the evolvement of trust during AI implementation as well as related organizational measures within the changing work environment. Hence, this research seeks to answer the following research question: *How do workers in changing work environments build trust in AI based systems, given the special role of AI reliability?*
Methodology

The analysis is based on a study within one of the biggest automotive suppliers. In this context, AI-based visual inspection was implemented into the running operations at more than 100 manufacturing lines in 7 international plants, in order to foster and further improve the high-quality standards. Ensuring a high-quality manufacturing process is crucial, as the goods are used for safety critical applications within vehicles. Thus, the evolution of trust during this embedded AI implementation represents the main research object.

Data collection: In order to gather a deep understanding of the ecosystem, we have involved ourselves deeply into the context. During data collection and analysis an iterative approach was followed by means of using grounded theory methodology (Glaser & Strauss, 1967). The sensitive research setting of embedded AI gave indication for selecting a qualitative approach (Li & Hahn, 2022). Primarily, data was collected using semi-structured interviews. To investigate the AI implementation from different perspectives, 20 workers of various organizational roles and from different hierarchical levels were selected. Beyond that, employees from differing plants and stages of implementation were considered. This allows to make comparisons along distinct phases of AI implementation and operation.

Data analysis: To obtain a profound overview of the key messages during the data collection process, we wrote a short summary with the most relevant statements after each interview (Glaser, 1978). Based on grounded theory, we used open, axial and selective coding for analyzing the interview transcripts (Corbin & Strauss, 2015). In a first step, open coding was used to identify general mechanisms that support the process of building trust in AI. However, during the data analysis process, we realized varied mechanisms at the different phases of AI implementation. That is the reason why we used axial coding to identify the respective phases and cluster the mechanisms accordingly. Moreover, we could figure out the expectations of the workers regarding AI reliability. In selective coding, we further developed and combined the dimensions, resulting in the final consolidation.

Preliminary results

The results show that the trust building process of workers in AI can be described according to five phases of AI implementation. Within each phase, organizational measures are revealed which support the implementation through five distinct targets: (i) Organizational setting, (ii) Input data, (iii) Performance, (iv) Promotion and (v) Robustness (Table 1).

Strikingly, the results reveal that these organizational measures not only enhance workers’ level of trust in AI, but also interrelate with workers’ expectations on AI reliability. At the beginning of the implementation, the workers seek for a very high degree of reliability, while – at the same point in time – there exists a low level of trust. This is caused by the fact that the workers are not willing to accept any deviation of the AI model with respect to reality which is also clearly communicated as a requirement to the AI development team. However, the results show that in the course of the implementations, the workers are willing
to deviate from that initial requirement which means that they accept a lower degree of reliability as compared to their original statement. Surprisingly, this reconfiguration has no negative effect on the building of trust, it is even the opposite. Various underlying organizational measures can be evaluated. As a conclusion, as one of the major key findings it can be stated that trust in AI exceeded to a such high extent that the workers engage with AI in a natural manner although the level of AI reliability had to be adjusted and lowered.

Table 1. Summary of the findings on AI implementation and its implications for workers

<table>
<thead>
<tr>
<th>Phase</th>
<th>Organizational measures</th>
<th>Reconfiguration of workers’ expectations on AI reliability</th>
<th>Effects on the level of trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Preparation</td>
<td>Organizational setting:</td>
<td>Workers request “100% reliability” of the AI model: They [the AI developers] told us: “Now we have [...] a slip through of 1, 2 or 3%. ” Then I said: “You can forget that. We demand 0% slip through from our machines and now we are suddenly supposed to allow 1%. That's not going to happen, it won’t work!” (#02)</td>
<td>Workers have a low level of trust in AI; first organizational measures prevent a trust level of zero.</td>
</tr>
<tr>
<td>#2 AI training</td>
<td>Input data:</td>
<td>Workers invent the label “unknown” for input data which they even cannot classify by themselves; documentation of “out of scope” parts, incl. the underlying reasons.</td>
<td>Workers acquire a basic understanding regarding the AI training process and the pitfalls in terms of data quality.</td>
</tr>
<tr>
<td>#3 AI verification</td>
<td>Performance:</td>
<td>Workers evaluate risk and continuously check AI performance of different groups (e.g. locations &amp; plants) on a profound dashboard; internal benchmarking starts the collective learning process: The AI classifies in the background, and you can check: What does the AI say? What does the operator say? Is there a mismatch? Then you can see whether the AI or the operator was wrong. (#03)</td>
<td>Workers scale AI, even with the constraint of a slightly lower level of reliability.</td>
</tr>
<tr>
<td>#4 AI rollout</td>
<td>Promotion:</td>
<td>Workers explain the new AI technology and its pitfalls towards outside parties; enhanced ability to engage with customers in case of occurring errors.</td>
<td>Workers trust AI, given certain boundary conditions.</td>
</tr>
<tr>
<td>#5 AI operation</td>
<td>Robustness:</td>
<td>Workers interact with AI in a natural and continuous way.</td>
<td>Workers trust AI and its continuous development.</td>
</tr>
</tbody>
</table>

Based on these findings a model for AI implementation is conceptualized (Figure 1). The model combines the organizational measures of the AI implementation phases with the
reconfigurations of workers expectations on AI reliability and the level of trust in AI.

Potential contributions

Our research builds on studies which investigated trust in technology (Lankton et al., 2015; McKnight et al., 2011; Söllner et al., 2016) in combination with studies specifically focusing on trust in AI (Glikson & Woolley, 2020; Li & Hahn, 2022). As a new perspective we focus on the dynamic process of building trust in AI, including the special role of workers’ expectations regarding AI reliability. We assign reliability this important part, as from our point of view the basic assumptions regarding the role of reliability for enhancing trust in technology do not hold in the context of non-transparent AI systems. The present work contributes towards prior discussions in the context of IT systems, emphasizing that a lower reliability will avoid the application of the technology by the workers (Lankton et al., 2015). We can give evidence that workers are willing to reconfigure their expected level of reliability in combination with their increasing experience with the AI system. This finding extends current studies which highlight the importance of a continuous AI implementation process (Hengstler et al., 2016). Further contributions can be made regarding the specific target of the measures within each phase. It can be concluded that the organizational measures within the third phase have the largest lever to build trust. We can confess that it is crucial to provide clarity on the AI performance in that phase. Prior studies showed that workers might not be capable to quantify uncertainties of AI models by themselves in an adequate way. It creates tension which might slow down or even stop the process of building trust into AI (Thiebes et al., 2021). We are planning to further concentrate on this and study the analogous patterns which help workers to reconfigure their expectations on AI, as well as the interconnection with organizational measures.

Figure 1. Dynamic process of AI implementation
References


