Also Available in English

What people really trust with AI

Still nothing new!?

Volume 4, Nr. 6

February 2023

Michael Leyer University of Marburg

Lilian Do Khac University of Marburg & Adesso SE

Layout & Design: Oliver Behn

White Paper Series of the Chair ABWL:
Digitalisation and Process Management
Volume 4

Trust with AI is one of the big topics these days and potentially one of the leading concepts that mediates between human and AI in the future. New empirical insights are created as well as new models and factors to identify the presence or absence of trust with AI. While there are lots of studies on the perception of AI, the question is what is really new in this regard. The underlying theories are known and most individuals do not understand what AI is when deciding to reject or accept to use services or to collaborate. And yet there are differences to traditional software that might not be consciously present.


Foto by macrovector / Freepik

Whom we trust with AI

The first aspect is the reference person/object of trust. With AI the main difference compared to traditional software is that it develops own strategies within boundaries set in order to solve the tasks given. AI learns from the data provided and it is often unclear how the results achieved are calculated respectively whether they are adequate. AI is however not objective but is driven by normative views incorporated either in the data provided, in the way the algorithm is trained, initially programmed or how it is applied. Hence, there are more actors than the AI to trust with. Next to the AI itself, this can be the owner of the AI, the programming company (if not the same), the training company (if not the same as the owner),the company using the AI to offer services to customers (if not the same as the owner) and the normative stance on AI (regulation and standards).

Who trusts AI

The question who is considering trust with an AI (or the companies employing it) is also of high importance as it determines their ability to assess an AI. Lay people who have no knowledge about AI won't be able to understand the efforts and infrastructure behind. Thus, they will not be able to take e.g. the training and programming into account. They might rather judge on characteristics of an AI itself or the provider using it. Results of an empirical study conducted show that indeed trust is in certain situations with the AI owner that then leads to overriding missing trust with an AI in order to use the AI. Moreover, for lay people it is often other features than the nature of the AI itself that are considered. This can be e.g. the voice being used by an AI which might be a popular actor with whom trust is associated with. Another example is using a popular character if an avatar is present. Such factors can override other aspects like results

quality of an AI. The nature of the algorithm is often not relevant as lay people do not understand the details even if they know which type of algorithm is applied. We showed in an empirical study that lay people develop similar trust, canniness and acceptance independent of the nature of an algorithm. Even if these algorithms are characterized by basic features that are opposite and would be expected to challenge different attitudes.

Contrary, experts, for example from a regulatory body, will however judge on the whole value chain including all actors and analyze them in detail. They will be interested in understanding which data is used, how it is processed and why an AI derives its results. In a similar manner, companies employing AI from other providers will be interested in such details to ensure that an AI is performing accordingly in order to achieve minimal risk conditions. If factors beyond the functional nature of AI are considered, these are typically related to how an AI is threatening or supporting a managers' job.

Understanding trust with AI

In order to understand why individuals in these different contexts have or develop trust with an AI or related actors, there is quite some theory available. One of the standard models describes that ability, benevolence and integrity lead to trust. Trust fueled by these antecedents can be dispositional, situational and learned. Further models adapted to AI described rather nuances of conceptual differences, but the basic underlying logic remains similar. Humans try to understand the behavior of an AI similar to understanding other humans. These also show certain characteristics and often remain black boxes. When assessing AI it is more difficult to use prior experiences learned with humans and the situation is often more ambiguous. And it is here that the main novelty is with AI – the specific parameters that are differently assessed between humans, machines and learning machines. Humans often try to make predictions on AI behavior derived from past experiences with humans, have assumptions

Picture by Freepik

from movies if it is clearly highlighted that an AI is acting or compare AI performance with human (their own) performance. Therefore, it is necessary to identify the specific parameters that are originating in the functional nature of AI, its appearance or behavior. The categories and relationships of the fundamental trust models inform the underlying logic of how trust is formed which follows the same patterns as with humans.

Experts will conduct such an assessment in a more analytic manner and more focus on the different aspects of data, algorithm and companies involved. In the end, a central assessment is made for an application while the assessment would be more scattered for a network of human actors performing the job. In the latter case it would also be necessary to determine human behavior with the different actors in order to assess trust with the services provided while the result of such behavior is engraved in the AI design and can be tested as such. This requires different procedures but also provides more potential of evidence to be tested analytically.

Conclusion

While the fundamental mechanism how trust evolves remains the same, the parameters of assessment can have a different relevance and priority between humans and AI. AI is missing quite some features that humans have in terms of appearance, but this represents different features in the details or thresholds to be considered within the existing mechanisms. This is where the novelty can be explored to understand why and when people from different backgrounds trust AI.

CONTACT DATA

Prof. Dr. Michael Leyer
Chair of ABWL:
Digitalisation & Process Management

Department Business and Economics

Adjunct Professor, School of Management,
Queensland University of Technology,
Brisbane, Australia

Email michael.leyer@wiwi.uni-marburg.de