Short Paper

Trust of (self-directed) learners in the use of artificial intelligence in an organizational context. A theoretical conceptualization.

Abstract

The interest of this paper is the trust of (self-directed) learners in the application of artificial intelligence (AI) in organizations. There is a relationship in decision making from the (self-directed) learner to the manager and to the AI. The guiding research question is: What are the critical variables of trust in the relationship between manager, used AI and user? The research method consists of a systematic literature review based on Tranfield et al. (2003). The result is a concept of trust based on the integrative model of organizational trust by Mayer et al. (1995), extended by aspects of initial trust by McKnight et al. (1998) and FEAS-elements of trustworthy AI by Toreini et al. (2020). This concept provides a starting point for further empirical studies.

Key Words: interpersonal Trust, organizational Trust, trustworthy Artificial Intelligence, trustworthy machine learning, Self-directed Learning

1. Introduction

The interest of this research is the trust of (self-directed) learners in the application of artificial intelligence (AI) in organizations. There is a relationship in decision making from the (self-directed) learner to the manager and to the AI.

AI is „a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Haenlein & Kaplan, 2019).

Global sales of enterprise applications in the AI sector are estimated to be around 4.8 billion US dollars in 2020, and 31.24 billion US dollars in 2025 (Statista Research Department, 2020). According to Statista, AI is one of the megatrends that is already changing our society and will probably continue to do so in the future (Breitkopf, 2020).

In the field of self-directed learning AI is already being used as adaptive learning systems or intelligent tutoring systems. AI is able to measure learner's competencies, suggest learning contents accordingly, determine learning styles, ask questions, and provide coaching if necessary (Goertz, 2014; He et al., 2019).

AI is controversially experienced and discussed. A survey of more than 8000 employees in 10 countries (He et al., 2019) showed that 64% of employees would trust a robot more and 50% would rather ask a robot for advice than their manager.

On the other hand almost 9 out of 10 companies stated in a survey of 1580 executives (Capgemini Research Institute (2019) that the use of AI in the organization led to ethical problems. Employees reported a disproportionate selection of candidates based on gender, ethnicity, age or other factors; the monitoring of employees in the workplace without their permission; and the misuse of personal data.

The theoretical and empirical research on interpersonal trust in an organizational context is comprehensive (Lewicki et al., 2006; Mayer et al., 1995; McKnight et al., 1998). In the area of trust in

AI, research deals rather with technical design options (Ezer, 2019) and the inclusion of ethical principles (Rossi, 2018; Toreini et al., 2020). The discussion of the perception of trustworthiness in AI from the perspective of the employees in an organization has not yet taken place (Toreini et al., 2020).

2. Objective

The aim is to develop a theoretical concept of critical trust variables in the relationship between the manager, the AI used and the user. The goals are to gain understanding of the trust variables and their effects, and to develop a theoretical basis for testing these variables. The guiding question is: *What are the critical variables of trust in the relationship between manager, used AI and user?*

3. Research Method

In order to capture concepts of trust in science, a systematic literature search was carried out using defined search strategies and criteria (Tranfield et al., 2003). Searches were conducted in EBSCO, EMERALD, SAGE Journals, Springer, researchgate.net, academia.edu and google scholar. In all databases the following search queries were made: trust in organizations, trust in artificial intelligence, trust in AI as learning support. The review is limited to German and English contributions. On this basis, key sources on theoretical models of trust in organizations, which have been empirically tested were identified (Tab. 1.1). Especially the Human Factors and Ergonomics Society provides contributions that deal with trust in human-AI teams in a military context (Tab. 1.2). Furthermore, there are studies on web-based and private uses of AI (Tab. 1.3). The search was iteratively extended with the keywords Machine Learning, interpersonal Trust, Technological Acceptance Model. On the basis of the reference to already existing trust constructs, two relevant contributions dealing with the trustworthiness of AI were found (Tab. 1.4). Inclusion criteria were: the definition of interpersonal trust in organizations; the distinction between initial and dynamic trust, trust outcomes; quantitative studies for measuring interpersonal trust in organizations; qualitative studies for measuring the perception of AI trustworthiness in organizations.

Table 1.1: Key sources for the development of the trust concept: theoretical models of trust in organizations (which have been empirically tested)

Study	Trust Components	Trust Object	Method
Lewicki et al. (1998)	Trust and distrust	Social Relationships	Conceptual
Lewicki et al. (2006)	Initial trust and trust development	Interpersonal trust	Conceptual
Mayer et al. (1995)	Trust proposition of trustor, factors of trustworthiness (ability, benevolence, and integrity = ABI) of trustee, risk-taking in the relationship, outcome	Interpersonal trust in organizational settings	Conceptual
Mayer & Davis (1999)	factors of trustworthiness (ABI)	Interpersonal trust in organizational settings	Empirical (quantitative)
McAllister (1995)	cognitive-based and affect-based trust	Interpersonal trust in organizational settings	Empirical (quantitative)
McKnight et al. (1998)	Concept of initial trust (including trust disposition, institutional trust, cognitive processes); trust definition adapted from Mayer et al. (1995)	Interpersonal trust in organizational settings	Conceptual

Table 1.2: Key sources for the differentiation from private consumers of AI and web-products			
Study	**Trust Components**	**Trust Object**	**Method**
Gefen et al. (2003)	Trust and TAM – technical acceptance model, Initial trust Variables in reference to McKnight et al. (1998)	consumer trust in e-vendor	Empirical (quantitative)
Li et al. (2008)	Initial trust Variables in reference to McKnight et al. (1998)	Trust in New Technology	Empirical (quantitative)
McKnight et al. (2002a)	Initial trust, Trusting beliefs dealing with ABI; Trusting intention as willingness to interact with an e-vendor	consumer trust in e-vendor	Empirical (quantitative)
McKnight et al. (2002b)	disposition to trust, institution-based trust, trusting beliefs, and trusting intentions	consumer trust in e-vendor	Empirical (quantitative)
Stewart (2003)	Initial trust	Trust on the World Wide Web	Empirical (experimental)
Table 1.3: Key sources for differentiation from the military context			
Study	**Trust Components**	**Trust Object**	**Method**
Ezer (2019)	Trust Engineering including trust outcomes	Human AI-Teams (military context)	Conceptual
Madhavan & Wiegmann (2004)	Comparison of human-automation teams with human-human partnerships	Human-automation teams (aviation context)	Conceptual
Sanders et al. (2011)	Model of Human-Robot Trust (Human, Environmental, Robot Characteristics)	Human-Robot-Interaction (military context)	Conceptual
Table 1.4: Key sources for trustworthiness of AI			
Study	**Trust Components**	**Trust Object**	**Method**
Siau et al. (2018)	Refers to ABI+predictability, trust intention, initial trust and trust development, HET (Human, Environmental, technical Characteristics)	Difference between Trust in AI and Trust in other Technologies	Conceptual
Toreini et al. (2020)	FEAS-technologies (Fair, Explainable, Auditable, Safe), with reference to ABI+, HET, initial and dynamic trust, and AI Principles	Trustworthiness of machine learning	Conceptual

Table 1: Overview of key results of the literature review

Findings: Theoretical Conceptualization of Trust in the relationship of the user of AI to the manager and to the used AI

In order to conceptualize trust in the context of management, used AI and users, it is necessary to use an integrative definition. Mayer et al. (1995) define trust as „the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action

important to the trustor, irrespective of the ability to monitor or control that other party“ (p. 712). The authors distinguish between characteristics of the trustworthy and the trust-giving person (p. 716).

Trustworthiness of manager. A manager is perceived trustworthy if he or she shows *ability, benevolence and integrity* (Mayer et al., 1995, p. 717).

Trust Disposition of User. This perception is influenced by the *propensity to trust* (Erikson, 1994; Rotter, 1967) of the trust-giving person. In their concept of initial trust, McKnight et al. (1998) add the element of the subjective decision of the trusting person (*trusting stance*) to make a leap of faith (p. 477). It can be argued that this trust can also be given to AI if the trusting person assumes that technical progress is good for mankind.

Institutional Trust of User. Another aspect is the influence of the organization. McKnight et al. (1998) refer to this as the trust in the institution. This consists of the belief in *structural security*, i.e. there are contracts, regulations or guarantees that establish a feeling of safety (Shapiro, 1987). This could include Basic Data Protection Regulation or (inter)national AI principles (Toreini et al., 2020, p. 280). A second aspect is the belief in *situational normality*. We could assume that in contexts where technical developments and products are normal, the use of AI is also considered normal and vice versa.

Cognitive Processes of User. Cognitive trust is a.o. based on first impressions (Lewis & Weigert 1985). These can be influenced by a *categorization*, like the reputation of the trusted person, group membership or stereotyping. Furthermore, McKnigth et al. (1998) describe the *illusion of process control*, a situational action of the trusting person towards the trusted person in order to establish initial trust. This can consist of a smile, and smiling back. It can be argued that the more a person engages with AI, the more they trust it. How sustainable the trust will then be will become apparent in further application.

Trustworthyness of AI. The implementation of technological solutions does not make the system trustworthy per se. Trustworthiness also requires basic technological qualities such as accuracy, efficiency and the performance of the algorithms (Siau et al., 2018; Toreini et al. 2020). There are also other elements, such as a user-friendly graphical user interface, which have an effect on trust (Davis, 1989). The FEAS technologies are specifically related to AI and therefore used here. FEAS stands for *Fair, Explainability, Auditability and Safe Technologies* (Toreini et al., 2020).

Trust Intention. The intention of trust consists in the *will to take a risk* and e.g. make data available and in the *will to depend on the person (or here AI) to be trusted* (McKnight et al., 1998).

Outcome. As a result, trusting actions or a trusting attitude of trustor would become visible. This includes: *Frequency of interaction with AI, Task handoffs, Information seeking behaviour, sharing information commitment, Accuracy of judgements on users, Distrust as a positive component, Organizational Commitment* (Ezer, 2019; McKnight, 2002a).

Integrating these findings following figure of the concept emerges (Fig. 1):

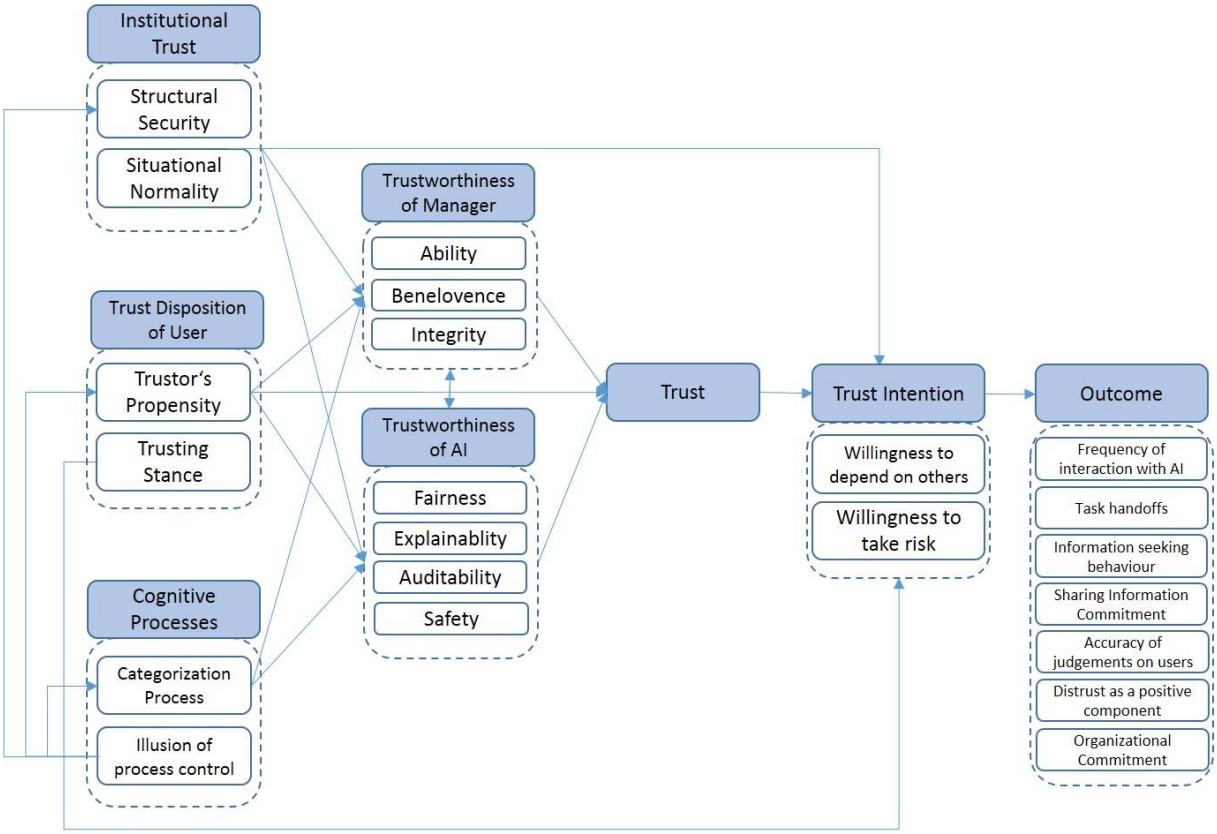

Figure 1: Integrative concept of trust in the relationship of the User of AI to the manager and to the used AI

4. Research Implications

The concept implies a qualitative approach as a first further step of empirical research. Thus, how signals of trustworthiness are perceived, can be identified and evaluated. In a further quantitative step, the trust variables could be tested. Various variables such as age and hierarchy can be considered here. The influence of different cultures is not viewed. Details need to be planned and discussed.

5. Practical Implications

Trust is an important factor in the successful implementation and acceptance of AI products in companies. The integrative approach contributes to more clarity in dealing with AI in the implementation process and in later stages of working with AI.

6. Originality

The originality consists in the comprehensive approach and the distinction between trustworthiness of the manager and trustworthiness of the AI used. Furthermore, the perspective is shifted away from a manager survey to an investigation of users (employees). In this respect, this study is the first to present such conceptualization.

7. Key References

Breitkopf, A. (2020). „Statistiken zu Megatrends“, Statista. Zugriff am 06.07.2020. Verfügbar unter https://de.statista.com/themen/3274/megatrends/

Capgemini Research Institute (2019). „Studie AI & Ethics. 9 von 10 Unternehmer sehen ethische Defizite bei Nutzung von künstlicher Intelligenz.“, Capgemini Research Institute. Verfügbar unter https://www.capgemini.com/at-de/news/studie-ethik-bei-kuenstlicher-intelligenz-entscheidend-fuer-vertrauen-in-unternehmen/

Colquitt, J. A.; Scott, B. A.; LePine, J. A. (2007). „Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships With Risk Taking and Job Performance.“ Journal of Applied Psychology, Vol. 92 (4), pp. 909-927

Davis, F.D. (1989). „Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” MIS Quarterly 13 (3), pp. 319-340.

Dietz, G.; Hartog, D. N. (2006). „Measuring Trust inside Organisazations.” Personnel review, Vol. 35 (5). pp. 557-588.

Dirks, K.T.; Ferrin, D.L. (2001). „The Role of Trust in Organizational Settings.“ Organizational Science, Vol. 12 (4), pp. 450-467

Dirks, K.T.; Ferrin, D.L. (2002). „Trust in Leadership: Meta-Analytic Findings and Implications for Research and Practice.“ Journal of Applied Psychology, Vol. 87 (4), pp. 611-628

Erikson, E.H. (1994). „Identity, youth and crisis.” Norton, New York

Ezer, N. (Hrsg.). (2019). „Trust Engineering for Human-AI Teams.“ Sage Journals. Proceedings of the Human Factors and Ergonomics Society Annual Meetings, 2019, Vol. 63(1), pp. 322-326.

Gefen, D. ; Karahanna, E. ; Straub, D.W. (2003). „Trust and TAM in Online Shopping: An Integrated Model.“ MIS Quarterly, Vol. 27(1), pp. 51-90

Goertz, L. (2014). „Digitales Lernen adaptiv Technische und didaktische Potenziale für die Weiterbildung der Zukunft.“ MMB-Institut für Medien- und Kompetenzforschung im Auftrag der Bertelsmann-Stiftung, Gütersloh 2014.

Haenlein, M.; Kaplan; M. (2019). „A Brief History of AI: On the Past, Present and Future of AI.“ California Management Review, 2019, Vol. 61 (4), pp. 5–14.

He, E.; Bertallee, C.; Jones, S.; Lyle, L.; Meister, J.; Schawbel, D. (2019). „Oracle & Future Workplace AI@Work Study 2019: From Fear to Enthusiasm: AI is Winning more Hearts and Mindes in the Workplace.“ AI@Work Study 2019. Zugriff am 06.07.2020. Verfügbar unter https://www.oracle.com/a/ocom/docs/applications/hcm/ai-at-work-ebook.pdf

Krasnova, H.; Spiekermann, S.; Koroleva, K.; Hildebrand, T. (2010). „Online social networks: why we disclose.“ Journal of Information Technology (2010) 25, pp. 109–125.

Lewicki, R.J.; McAllister, D.J.; Bies, R.J. (1998). „Trust and Distrust: New Relationships and Realities.“ The Academy of Management Review. Vol. 23 (3), pp. 438-458

Lewicki, R. J.; Tomlinson, E. C.; Gillespie, N. (2006). „Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions.“ Journal of Management, Vol. 32 (6), pp. 991-1022.

Lewis, J. D.; Weigert, A. J. (1985). „Social atomism, holism, and trust.” The Sociological Quarterly, Vol. 26, pp. 455-471

Madhavan, P.; Wiegmann, D.A. (2004). „A New Look At The Dynamics Of Human-Automation Trust: Is Trust In Humans Comparable To Trust In Machines?“ Proceedings Of The Human Factors And Ergonomics Society 48th Annual Meeting, pp. 581-585.

Mayer, R. C.; Davis, J. H.; Schoorman, F. D. (1995). „An Integrative Model of Organizational Trust.“ Academy of Management Review, 20 (3), pp. 709-734.

Mayer, R. C.; Davis (1999). „The Effect of Performance Appraisal System on Trust for Management: A Field Quasi Experiment.“ Journal of Applied Psychology, Vol. 84 (1), pp. 123-136.

McAllister, D. J. (1995). „Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations.“ Academy of Management Journal, Vol 38 (1), pp. 24-59

McKnight, D. H.; Cummings, L. L.; Chervany, N. L. (1998). „Initial Trust Formation in New Organizational Relationships.“ Academy of Management Review, 23 (3), pp. 473-490.

McKnight, D. H.; Choudhury, V.; Kacmar, C. (2002a). „The impact of initial consumer trust on intentions to transact with a web site: a trustbuilding model.“ Journal of Strategic Information Systems, Vol. 11, pp. 297-323

McKnight, D. H.; Choudhury, V.; Kacmar, C. (2002b). „Developing and Validating Trust Measures for e-Commerce: An Integrative Typology.“ Information Systems Research, Vol. 13 (3), pp. 334-359

Rossi, F. (2018). „Building Trust in Artificial Intelligence.“ Journal of International Affairs, 72 (1), pp. 127-134.

Rotter, J.B. (1967). “A new scale for the measurement of interpersonal trust.” Journal of Personality, Vol. 35 (4), pp. 651–665.

Sanders, T.; Oleson, K. E.; Billings, D. R.; Chen, J. Y. C.; Hancock, P. A. (2011). „A Model of Human-Robot Trust: Theoretical Model Development.“ Sage Journals. Proceedings of the Human Factors and Ergonomics Society Annual Meetings, 2011, Vol. 54, pp. 1432-1436.

Shapiro, S. P. (1987). „The social control of impersonal trust.” American Journal of Sociology. Vol. 93, pp. 623-658

Siau, K.; Wang, W. (2018). „Building Trust in Artificial Intelligence, Machine Learning, and Robotics.“ Cutter Business Technology Journal, Vol. 31 (2), pp. 47-53.

Statista Research Department (2020). „Prognose zum Umsatz mit Unternehmensanwendungen im Bereich künstliche Intelligenz in Europa von 2016 bis 2025.“, Statista. Zugriff am 06.07.2020. Verfügbar unter https://de.statista.com/statistik/daten/studie/620513/umfrage/umsatz-mit-anwendungen-im-bereich-kuenstliche-intelligenz-in-europa/

Stewart, K. J. (2003). „Trust transfer on the World Wide Web.“ Organization Science, Vol. 14 (1), pp. 5-17

Toreini, E.; Aitken, M.; Coopamootoo, K.; Elliott, K.; Zelaya, C. G.; van Moorsel, A. (2020). „The relationship between trust in AI and trustworthy machine learning technologies.“ Proceedings of Conference of Fairness, Accountability, and Transparency (FAT* ’20, January 27–30, 2020, Barcelona, Spain) of Association for Computing Machinery (ACM), pp. 272-283

Tranfield, D.; Denyer, D.; Smart, P. (2003). „Towards a methodology for developing evidence-informed management knowledge by means of systematic review.“ British Journal of Management, Vol. 14, pp. 207–222.