Measuring Trustworthiness of AI Systems: A Holistic Maturity Model

Myriam Schaschek
Julius-Maximilians-University, myriam.schaschek@uni-wuerzburg.de

Sarah Engel
Philipps University, sarah@engel.link

Measuring Trustworthiness of AI Systems: A Holistic Maturity Model

Short Paper

Myriam Schaschek
Julius-Maximilians-University
Würzburg, Germany
myriam.schaschek@uni-wuerzburg.de

Sarah Engel
Philipps University Marburg
Marburg, Germany
engelsar@uni-marburg.de

Abstract

Artificial intelligence (AI) has an impact on business and society at large while posing challenges and risks. For AI adoption, trustworthiness is paramount, yet there appears to be a gap between theory and practice. Organizations need guidance in quantitatively assessing and improving the trustworthiness of AI systems. To address such challenges, maturity models have shown to be a valuable instrument. However, recent AI maturity models address trustworthiness only at the maturest level. As a response, we propose a model to integrate the concept of trustworthiness across the AI lifecycle management. In doing so, we follow Design Science Research to develop a holistic model highlighting the importance of trustworthiness throughout the AI adoption journey to realize the real value potential. This research-in-progress contributes to the emerging research on human-AI systems and managing AI. Our objective is to use the model for assessing, evaluating, and improving trustworthy AI on an organizational level.

Keywords: Trustworthy AI, AI maturity, AI adoption, managing AI, trustworthiness

Introduction

Trustworthiness and ethics of artificial intelligence (AI) are widely and controversially discussed in public, politics, and organizations (Hevner and Storey 2022; Robert Jr et al. 2020). While there are various frameworks that offer guidance for managing AI in organizations (Sadiq et al. 2021), the knowledge of risks and prevention of potential harm of AI systems is far from conclusive. For most organizations, concerns about the potential harm of AI systems are a barrier to its adoption in production (Someh et al. 2020). Empirical investigations show that relatively few AI pilot projects go to production and, thus, unfold low return on investment and business value (Benbya et al. 2020). Additionally, upcoming AI regulations, such as the EU AI ACT, will play an important role for organizations active in the EU to address compliance and trustworthiness in AI. Organizations have to act, at least for AI systems classified as high risk (Smuha 2019). Today's managers must deal with both the possibilities and risks accompanying AI.

The majority of organizations recognize the importance of trustworthy AI (TAI), but only a few organizations have taken action (IBM Corporation 2022). The need for TAI guidance becomes apparent as there exists a significant knowledge gap in managing TAI in practice (Morley et al. 2020; Robert Jr et al. 2020). In order to close this intention-action gap, organizations, in particular managers, need support to assess, evaluate and improve TAI (Jantunen et al. 2021; Robert Jr et al. 2020). Additionally, the adoption of TAI practices demands organizations to be prepared. Maturity models are action-oriented tools that contain discrete levels of organizational maturity along different organizational capability dimensions. Beyond representing the levels, maturity models address the management issues that arise on the adoption journey (Uren and Edwards 2023). This helps managers to evaluate and progress in defined focus areas. Accordingly, maturity models serve as a high-level roadmap for navigating forward in the endeavor.

Existing AI maturity models (AI-MM) indicate TAI only at the maturest level or neglect it (e.g., Alsheiabni et al. 2019; Fukas et al. 2021). Since most AI maturity models focus on the technical aspects of designing AI systems (Sadiq et al. 2021), a trade-off arises between implementing a model to demonstrate value or complicating the technical requirements with ethical considerations (Jantunen et al. 2021). However, we see the need to inform managers about the necessary steps towards trustworthy AI system design and aim to advance the understanding of the organizational journey towards AI adoption that includes TAI capabilities from early on.

So far, principles and value-based guidelines have been the primary focus in academia to provide guidance for organizations (Fischer and Beimborn 2022; Mayer et al. 2021; Rothenberger et al. 2019; Shneiderman 2020). Guidelines alone will not provide sufficient support for managers of organizations to adopt and use a TAI system. Managers require knowledge on the organizational capabilities needed to thrive on the adoption journey (Uren and Edwards 2023). In practice, it could pave the way to TAI systems by defining the vital organizational capabilities that are required while remaining aware of the technical specifics of AI. From these objectives, the following research questions (RQ) arise:

RQ1: How can a trustworthy AI maturity model be designed?

RQ2: What are the key organizational capabilities for building trustworthy AI?

In this *research-in-progress* paper, we apply Design Science Research (DSR) methodology to propose a preliminary trustworthy AI maturity model (TAI-MM) that meets guidelines and principles for TAI to support managers on their way to becoming a trustworthy AI-enabled organization. In doing so, we aim to create a model that helps to benchmark and promote proper AI capabilities and integrates TAI requirements across the AI lifecycle. With the design and evaluation of the preliminary TAI-MM, we significantly contribute to the information systems (IS) research community, filling a research gap and providing actionable guidance for bridging theory and practice. In particular, we contribute to the discourse on managing AI in organizations (Berente et al. 2021) and on human-AI interaction (Amershi et al. 2019; Hevner and Storey 2022).

Our paper structures as follows: In Section 2, we describe the theoretical background and related work. Then, section 3 covers the status quo of the TAI-MM design and evaluation strategy. Finally, we present preliminary results in Section 4 and finish this *research-in-progress* paper with a conclusion in Section 5.

Theoretical Background and Related Work

Towards Trustworthy Artificial Intelligence Systems

For the purpose of this paper, we consider Wang's definition of AI because it fits well for AI systems in the workplace (Wang 2019): "Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources." We emphasize the characteristics of AI systems to process (often a high amount of) information and learn over time, aiming to solve complex real-world problems that have not been possible with traditional rule-based approaches. With acceleration, AI systems are permeating the daily lives of society.

In response to ethical questions raised by applying AI in real-world scenarios, the European Union (EU) defined TAI systems as AI systems that fulfill the following requirements: respect for human autonomy, prevention of harm, fairness, and explicability (Smuha 2019). Starting from this, AI researchers and practitioners developed and proposed various TAI definitions, dimensions, guidelines, and interrelations (Thiebes et al. 2021). Despite the differences in terminology and content, the idea of TAI systems builds on the complex phenomenon of trustworthiness. It aims at advancing the arbitrary AI frontier to maximize its benefits while mitigating potential harm (Lockey et al. 2021). The concept of trust involves two sides: the party to be trusted (trustee) and the trusting party (trustor). We consider the notion of trustworthiness as the attribute of an AI system (as trustee) that the end-user (as trustor) must trust as an antecedent of the adoption and use of AI-enabled tools (Zhu et al. 2021).

Who is the actual user of trustworthy AI? The user can be the end customer, but in an organizational context, there are different possible types of users, e.g., engineers who are responsible for developing and adjusting performance, domain experts who get augmented or even replaced in their work (Hafermalz and Huysman 2021), auditors or regulators who assess and control AI systems, and managers who are held accountable for AI systems adoption and possible failures. However, guidance towards TAI systems primarily addresses AI system developers or end-users (Meske et al. 2020) while neglecting the organizational perspective. An organizational perspective on AI trustworthiness is necessary to advance TAI system design. Consequently, with our TAI-MM, we target primarily managers and explore organizational capabilities needed to create value from AI trustworthiness.

The Role of Organizational Culture in Adopting AI Systems

In discussing AI adoption, many studies mention the importance of organizational level and resort to the well-known theories Technology Organization Environment (TOE) and Socio-Technical Systems Theory (STS) (Jeyaraj et al. 2006; Alsheiabni et al. 2019; Uren and Edwards 2023; Yu et al. 2023). These theories help organizations in better understanding different factors that influence technology adoption and emphasize that the development and adoption of AI systems are intertwined with organizations' context and environment, including culture and processes. To analyze and systematize the organizational capabilities, we consider those theories as kernel theories to guide our maturity model design.

In the context of AI system adoption, the interdisciplinary concept of trustworthiness plays an important role that is well-explored in IS research (Mcknight et al. 2011; McKnight et al. 2002; Rousseau et al. 1998). Even though, previous research on trustworthiness and adoption of AI systems hardly covers the organizational context. Trustworthiness of AI systems was mainly investigated on a technology and application level (Thiebes et al. 2021, Shneiderman 2020, Lockey et al. 2021). And the currently discussed draft for EU regulation for AI systems demands organizations to act and motivates research ambitions. Our research-in-progress article wants to fill this gap and provides guidance for managing AI on an organizational scope. Thereon, we base on well-established approaches for technology adoption and trustworthiness.

Current (AI) Maturity Models in Information Systems Research

Now, how to shed light on organizational aspects of trustworthy AI and help managers navigate? Maturity models guide managers in balancing divergent objectives while assessing an organization's current capabilities. Over the years, a plethora of maturity models have been developed and applied in the field of IS research to assist in continuous improvement initiatives or to help organizations in terms of developing the organization by self-assessment, benchmarking and guidance. Maturity models originate from software engineering, where one of the most well-known models have been invented - the capability maturity model (Paulk et al. 1993). Due to the plethora of proposed maturity models, researchers investigated their characteristics and developed a classification system for IS maturity models (Mettler et al. 2010). The common feature of maturity models is that a certain number of dimensions are described at different maturity levels. Thereby, "maturity" is defined as "the state of being complete, perfect or ready" (Weiner and Simpson 1989). The need for maturity models has not yet abated, and models have been developed to capture the organizational perspective on technology-driven change (Felch et al. 2019).

Different maturity models have already been proposed in the context of AI system development. Two general directions can be distinguished in AI-MM research: the development of maturity models to assess the readiness level of AI systems in organizations in general and the development of a maturity model for specific domains. Examples are the AI-RFX Procurement Framework, AI Management Framework (Lichtenthaler 2020), or the AI-MM Framework (Ellefsen et al. 2019). Due to the multifaceted effects of AI systems on organizations, various perspectives on maturity span the field of AI system development. Thereby, the AI-MMs generally focus on the organizations' perspective and do not consider the specifics of ethical concerns and demand for TAI. Only on the most mature level the concept of trustworthiness is integrated. With this in mind, first attempts in research have been made towards assessing ethics-related AI systems development maturity (Jantunen et al. 2021).

As we aim to understand what are the capabilities needed for building TAI in organizations, we focus on AI-MM models that extend their focus on the ethical principles related to AI systems. In total, we identified only one research paper that targets AI-MM and ethical principles. Summarizing our findings, research investigating how organizations assess and benchmark AI maturity concerning the socio-technical concept of trustworthiness is still relatively nascent and insufficiently investigated. We will advance existing AI-MMs by introducing TAI on all maturity levels.

Research Design

Methodology Overview

DSR serves as the overarching methodology for the iterative model development process. Our research in progress adapts DSR (Hevner et al. 2004) and investigates the gap between the current and target state of TAI adoption. We propose the following research process to design a viable artifact for TAI adoption. Figure 1 illustrates the current research-in-progress activities and highlights our planned upcoming research design (grey background).

Following the DSR methodology by Peffers et al. (2007), our research process includes problem identification and motivation, objectives of a solution, design and development, demonstration, evaluation, and communication. At the present stage, we completed our anticipated artifact's first DSR process iteration. Thus, we are transitioning to phase *demonstration* after communicating our recent *research-in-progress* results.

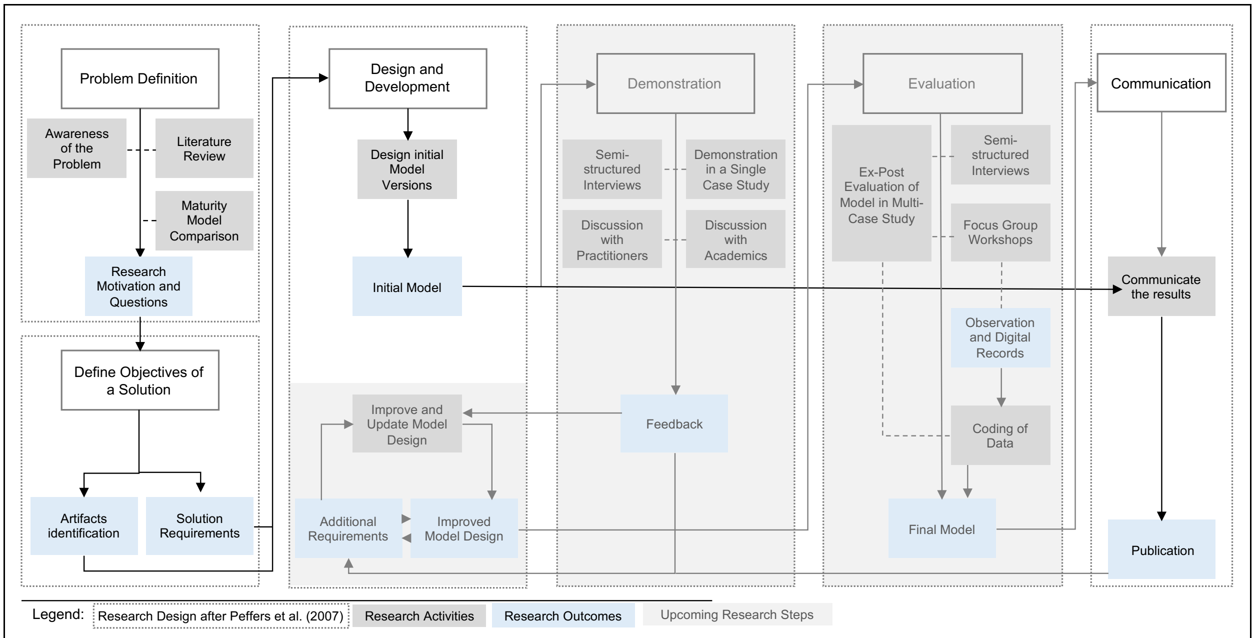

Figure 1. Our research design in accordance with Peffers et al., (2007)

Within the first DSR process iteration, we developed an initial design of our artifact, the TAI-MM. Our envisioned artifact aims to effectively solve the problem of AI adoption in organizations by providing guidance with our TAI-MM model. In the wake of our initial artifact design, we conducted a structured literature review (Brocke et al. 2009) regarding TAI capabilities and AI-MMs. Then, we drew on the descriptive and prescriptive knowledge of maturity model development and advances in TAI research, following vom Brocke et al. (2020). After defining the solution objectives and requirements to infer viable artifacts, we entered the *design and development* phase and created an initial artifactual solution following the iterative maturity model development process by Becker et al. (2009).

Literature Review

To identify the state-of-the-art AI-MMs and literature on TAI, we conducted an interdisciplinary and structured literature review following the guidance of vom Brocke et al. (2009). First, we identified the most relevant databases related to our research focus: Web of Science, IEEExplore, ACM, EBSCOhost, and AISeL. Second, we organized our review by focusing on central concepts related to TAI adoption in organizations. For the search, we used the keywords: (*Artificial Intelligence, AI*) AND (*trust*, trustworthy AI, TAI*) OR (*matur*, maturity model*). The complete literature search process is depicted in Figure 2, and the main findings that informed our initial artifact design are summarized in Section Theoretical

Background and Related Work. Finally, we identified a total of n=33 relevant publications, including a forward and backward search.

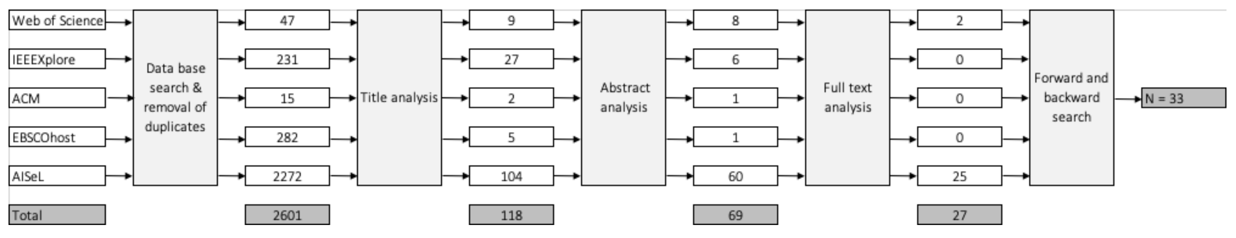

Figure 2. Our literature search process following vom Brocke et al. (2009)

Towards a Trustworthy AI Maturity Model

During the *design and development* of the initial maturity model for TAI, we used a grounded theory approach based on a structured literature review to qualitatively identify useful concepts for model building (Wolfswinkel et al. 2013). In doing so, we leveraged the codes identified in the structured literature and synthesized them into elements and model dimensions, see Figure 3. The elements of our maturity model are the TAI capabilities. Model dimensions build the columns. Rows are defining the maturity levels towards adoption of TAI. In a third step, the identified capabilities are assigned to the organizational maturity level.

Trustworthy AI Capabilities and Dimensions

Conforming to grounded theory, our procedure was guided by three coding stages: (1) open coding, (2) axial coding, and (3) selective coding (Glaser and Strauss 2017; Strauss and Corbin 1994). Figure 3 shows the coding scheme, including exemplary extracts from the coding. During *open coding*, we reread the core literature defined through the structured literature review and identified organizational capabilities for building trustworthiness of AI systems and adopting AI. Following *axial coding*, we grouped the identified codes into categories of capabilities for TAI systems. In the third step, we conducted *selective coding*, connected categories, and derived at TAI dimensions. After selective coding we matched our TAI dimensions to the socio-technical AI dimensions that demonstrated valuable for AI adoption. These dimensions are grounded on the TOE (Uren and Edwards (2023) and STS theories, which function as kernel theories for our artifact design proposal.

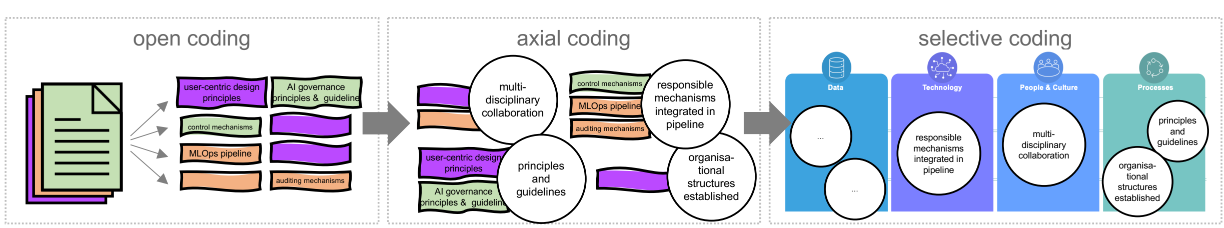

Figure 3. Coding scheme for trustworthy AI capabilities and dimensions incl. examples

Trustworthy AI Maturity Levels

Maturity levels represent discrete stages of TAI adoption. Given the design of the TAI-MM, identified TAI capabilities were brought into a logical, incremental order, interdependencies made transparent, and assigned to various maturity levels. The selection of the five levels (1-5) followed existing AI maturity models (e.g., Jantunen et al. 2021; Lichtenthaler 2020; Sadiq et al. 2021). In the first iteration, the assignment followed observations from real-world experience (vom Brocke et al. 2020).

Evaluation Strategy

The FEDS framework informs our artifact design to demonstrate its proof of utility (Venable et al. 2016). Due to the social context of our artifact, we decided to follow a two-staged human risk & effectiveness evaluation strategy (Venable et al. 2016) with the primary evaluation goal of rigorously validating the artifact's utility in real-world situations. By evaluating, we ultimately seek to demonstrate that the trustworthiness maturity assessment reflects the TAI-MM levels. During the first design cycle, we conducted a formative evaluation to reduce uncertainty in the model design. In this *research-in-progress* paper, we performed an ex-ante formative artificial evaluation by discussing the model against competing artifacts (Siau and Rossi 1998) and compared it with selected existing methods (see Section 2).

Upcoming Research: Second and Third Design Cycle

In upcoming research, we plan to perform ex-ante formative naturalistic evaluation activities, including semi-structured expert interviews and focus group workshops in the context of a real-world case study. To further validate whether our artifact design provides utility, effectiveness, and comprehensibility in different organizational contexts, we plan to conduct a multi-case study as ex-post naturalistic evaluation.

Case Study Design

Moving forward, we conduct a multiple case study to demonstrate the proof of utility of our developed model and extend the model with practical insights. In particular, a case study allows qualitative in-depth, multi-faceted explorations in complex real-world organizational settings. We follow Yin (2003) for the case study design. For the first case study, we have chosen a reputable global technology (incl. AI) company and will recruit at least eight experts. Professionals, who hold accountability for AI systems and have professional experience in developing TAI systems or managing AI, are considered TAI experts, e.g., Chief Privacy Officers, AI Ethics Board Members, and AI Model Owners. To gather feedback on the model's utility, TAI capabilities, and their assignment to maturity levels, we plan to conduct semi-structured interviews with named experts; and record and transcript interviews for data collection.

In general, multiple case studies are a fruitful methodological approach to deepen the understanding of AI trustworthiness across organizations. Thus, we plan to gain holistic and contextualized insights into adopting AI systems over various organizational ecosystems, stakeholders, and maturity levels. This is likely to help us systematize and generalize the understanding of how organizations cope with our developed TAI-MM. Specifically, we plan to present the TAI-MM at a TAI Summit where multiple (more than 20) managerial stakeholders from different organizations and industries attend. Experts are asked to self-assess their organizations' TAI maturity level based on the presented TAI-MM. Based on the assessed level of maturity, we plan moderated discussions in smaller groups about needed capabilities for their organization against the TAI capabilities proposed in the TAI-MM.

Preliminary Trustworthy AI Maturity Model

After completing the first iteration of the TAI-MM development, we derived four AI dimensions and defined five progressive maturity levels. Figure 4 displays the preliminary TAI-MM. With this, we answer RQ1 by providing extensive and detailed information on how a TAI-MM can be designed. Every maturity level has been assigned based on real-world observations and our extensive literature review on TAI capabilities.

While coding our TAI dimensions, we found a match to the socio-technical AI dimensions proposed by Uren and Edwards (2023). This allows us to extend existing AI-MMs with our identified TAI capabilities without introducing new dimensions. The preliminary TAI dimensions are Data, Technology, People & Culture, and Processes. Data builds a solid foundation for AI, and organizations need capabilities for trustworthy data collection, processing, and generation. Data quality assessment, bias mitigation, data policy, data governance, data strategy are key capabilities. The second dimension is technology, where capabilities around metrics for the trustworthiness of AI products and services, architecture, tools, and the AI lifecycle are important. Third, people and culture elaborate on organizational capabilities in education, staffing, resources, top management support, advocacy network, team setup, communication, and ways of working. The fourth dimension covers the processes and structures in an organization, incl. principles, guidelines, policies, roles, and responsibilities.

Organizations that just get started are aware of AI systems' potentials, limitations, and risks. In level 2, initial steps are taken to build skills and explore and define principles and accountability. Moving to the third level, organizations now approach AI systems strategically, know about their weak points and address them holistically. Next, level 4 is indicated by operationalization and the ability to scale the organization's processes, data & technology practices, and infrastructure and live up to their values across the organization. Finally, the maturest level is characterized by an infusion of trustworthy and ethical standards throughout business strategy, leadership, culture, ecosystem, data & technology practices, and infrastructure underpinning trustworthiness and performance. This is answering RQ2 by providing this first *research-in-progress* version. Further investigations, evaluation, and refinement are the scope of further research.

	Data	Technology	People & Culture	Processes
Level 1 Being aware	• Extensive Data Collection with low Privacy & Security standards • **Awareness** about impact of Data on AI and output quality	• Attitude towards fairness, explainability, accountability, sustainability, transparency • **Awareness** about constraints related to current technology	• **Understanding** of AI, intention, ambition, risks and limitations • **Attitude towards AI ethics**	• **Awareness** of the need for organizational processes and transparency for them
Level 2 Taking first steps	• Data Governance **Skills** onboarded (trained / hired / Consultants) • Ad Hoc **Attempts** towards Data quality (efforts to improve data quality), Data processing and Data understanding (efforts to understanding organization's data & algorithms)	• **Exploration & Prototyping** - Use of forensic tech for detection & mitigation (explainable, fair, transparent, accountable)	• **Skills & Education** • **Cross-disciplinary** debate • Engagement with partners from academia, Think Tanks or participation at events	• **Principles & guidelines** for guidance defined & communicated internally & externally • New role for CRaiO/ Ethics Leader/ CPO or Advisory Board announced • **Roles & responsibilities** defined (accountability)
Level 3 Approaching strategically	• **Data strategy:** actionable strategy covering data collection, quality and security • **Data governance** to fulfill GDPR requirements (responsible collection, processing, information security, data protection audits)	• Definition of **metrics** for trustworthiness of AI products and services - defined with multidisciplinary & diverse stakeholders • Definition of **technical Framework & architecture** to enhance trust • **Communication**: information about algorithms, human-AI interaction and decision-making incl. error	• **Resources** available: experts, trainings, budget • **Participation** in cross-industry, government and scientific initiatives and events • Multidisciplinary effort: employee involvement and cross-disciplinary collaboration • **Communication** strategy defined	• Awareness & **Holistic assessment** of weaknesses: impact & risks, culture & structures, from design onwards • **Strategy** includes socio-economic aspects: corporate standards, use of AI, etc. • **Processes** for risk mgmt., monitoring, auditing defined and partially implemented
Level 4 Operationalizing	• **Data policy** is transparent and customer-centric • **Data governance** streamlined across the organization, automated & integrated into processes • Data processing to actively mitigate bias	• **Responsible Design Practices** part of every AI project: responsibility by design, explainability understandable, fairness, simplest possible solution, human oversight, testing • **Tools integrated** into existing landscape and operational • Responsible mechanisms integrated & automated along **AI lifecycle**: control mechanisms, automated, MLOps - CI/CD, validation, monitoring, handy documentation	• Top management **commitment** & support • **Advocacy** network and communities who drive socially responsible initiatives • **Diverse teams** actionable and empowered to create trustworthy AI systems • Stakeholder engagement across disciplines and along the AI lifecycle • Communication strategy implemented and change management active	• **Organizational structures** (compliance departments, CRaiO, Advisory Board) have power and accountable for fallback scenarios • **Policy** established to ensure technology aligned with org's values • **Processes** for risk mgmt., fallback & failure, monitoring, auditing, approval or formal discussions of AI projects - adopted across the organization • Incentive mechanisms in place
Level 5 Innovating	• **Data exploration**: organizations can uncover problems by exploring data • **Holistic practices**: governance, ethical design, risk control etc.	• **Continuous improvement** & performance measure against Trustworthiness metrics • Appropriate architecture and infrastructure underpinning trust & performance • Continuous exploration, review and adoption of state-of-the art insights & trends	• Ethics seen and leveraged as **competitive advantage** across the organization, part of org's DNA • **Thought Leadership** in TAI • Externally recognized as top responsible employer for talents and trustworthy company • High **adoption rate** from employees and customer	• **AI Governance embedded** in organizational fabric • Broader **ecosystem** established and leveraged to innovate

Figure 4. Preliminary trustworthy AI maturity model

Conclusion

This *research-in-progress* paper presents the first steps towards developing a TAI-MM. We demonstrate preliminary results on vital organizational capabilities needed to create TAI systems. In doing so, our initial model proposes a holistic perspective on maturity levels across data, technology, people & culture, and processes. It highlights the importance of trustworthiness throughout the AI adoption journey. The superordinary benefit of our maturity model is the assessment, evaluation, and improvement of TAI on an organizational level. At the same time, we are aware of some limitations. The preliminary results mainly ground on literature data and observations from practice. Thus, future research needs to qualitatively evaluate whether the model design holds across organizations. To showcase the utility of the TAI-MM, we have spoken to managers and professional experts in the field of AI governance. These validated the need for such guidance and inspired further research endeavors. Next, we will follow our evaluation strategy and conduct multiple case studies to enhance our preliminary model design. In this vein, we contribute to the emerging research on human-AI systems and managing AI.

References

Alsheiabni, Sulaiman, Yen Cheung, and Chris Messom. 2019. "Towards an Artificial Intelligence Maturity Model: From Science Fiction to Business Facts." in *PACIS 2019 Proceedings*.

Amershi, Saleema, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, and others. 2019. "Guidelines for Human-AI Interaction." Pp. 1–13 in *Proceedings of the 2019 conference on human factors in computing systems*.

Becker, Jörg, Ralf Knackstedt, and Jens Pöppelbuß. 2009. "Developing Maturity Models for IT Management." *Business \& Information Systems Engineering* 1(3):213–22.

Benbya, Hind, Thomas H. Davenport, and Stella Pachidi. 2020. "Artificial Intelligence in Organizations: Current State and Future Opportunities." *MIS Quarterly Executive* 19(4):ix--xxi.

Berente, Nicholas, Bin Gu, Jan Recker, and Radhika Santhanam. 2021. "Managing Artificial Intelligence." *MIS Quarterly* 45(3):1433–50.

Brocke, Jan vom, Alexander Simons, Bjoern Niehaves, Bjorn Niehaves, Kai Reimer, Ralf Plattfaut, and Anne Cleven. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process." P. 161 in *ECIS 2009 Proceedings*.

vom Brocke, Jan, Robert Winter, Alan Hevner, and Alexander Maedche. 2020. "Special Issue Editorial--Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey through Time and Space." *Journal of the Association for Information Systems* 21(3):9.

Ellefsen, Anna Paula Tanajura, Joanna Oleśków-Szłapka, Grzegorz Pawłowski, and Adrianna Toboła. 2019. "Striving for Excellence in AI Implementation: AI Maturity Model Framework and Preliminary Research Results." *LogForum* 15(3).

Felch, Vanessa, Bjoern Asdecker, and Eric Sucky. 2019. "Maturity Models in the Age of Industry 4.0--Do the Available Models Correspond to the Needs of Business Practice?" Pp. 5165–74 in *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Feuerriegel, Stefan, Mateusz Dolata, and Gerhard Schwabe. 2020. "Fair AI." *Business \& Information Systems Engineering* 62(4):379–84.

Fischer, Vanessa, and Daniel Beimborn. 2022. "How Should Organizations Manage Artificial Intelligence? A Strategic Literature Review." in *PACIS 2022 Proceedings*.

Fukas, Philipp, Jonas Rebstadt, Florian Remark, and Oliver Thomas. 2021. "Developing an Artificial Intelligence Maturity Model for Auditing." in *ECIS 2021 Research Papers*.

Glaser, Barney G., and Anselm L. Strauss. 2017. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Routledge.

Hafermalz, Ella, and Huysman, Marleen. 2022. "Please explain: Key questions for explainable AI research from an organizational perspective." *Morals & Machines* 1.2: 10-23.

Hevner, Alan R., Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. "Design Science in Information Systems Research." *MIS Quarterly: Management Information Systems* 28(1):75–105.

Hevner, Alan R., and Veda C. Storey. 2022. "Research Challenges for the Design of Human-Artificial Intelligence Systems (HAIS)." *ACM Transactions on Management Information Systems*.

IBM Corporation. 2022. *AI Ethics in Action - An Enterprise Guide to Progressing Trustworthy AI*.

Jantunen, Marianna, Erika Halme, Ville Vakkuri, Kai-Kristian Kemell, Rebekah Rebekah, Tommi Mikkonen, Anh Nguyen Duc, and Pekka Abrahamsson. 2021. "Building a Maturity Model for Developing Ethically Aligned AI Systems." in *IRIS*.

Jeyaraj, Anand, Joseph W. Rottman, and Mary C. Lacity. 2006. "A Review of the Predictors, Linkages, and Biases in IT Innovation Adoption Research." *Journal of Information Technology* 21(1):1–23.

Lichtenthaler, Ulrich. 2020. "Five Maturity Levels of Managing AI: From Isolated Ignorance to Integrated Intelligence." *Journal of Innovation Management* 8(1):39–50.

Lockey, Steven, Nicole Gillespie, Daniel Holm, and Ida Asadi Someh. 2021. "A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions." in *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Mayer, Anne-Sophie, Alexandra Haimerl, Franz Strich, and Marina Fiedler. 2021. "How Corporations Encourage the Implementation of AI Ethics." in *ECIS 2021 Research Papers*.

Mcknight, D. Harrison, Michelle Carter, Jason Bennett Thatcher, and Paul F. Clay. 2011. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems (TMIS)* 2(2):1–25.

McKnight, D. Harrison, Vivek Choudhury, and Charles Kacmar. 2002. "Developing and Validating Trust

Measures for E-Commerce: An Integrative Typology." *Information Systems Research* 13(3):334–59.
Meske, Christian, Enrico Bunde, Johannes Schneider, and Martin Gersch. 2020. "Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities." 39(1):53–63.
Mettler, Tobias, Peter Rohner, and Robert Winter. 2010. "Towards a Classification of Maturity Models in Information Systems." Pp. 333–40 in *Management of the interconnected world*. Springer.
Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices." *Science and Engineering Ethics* 26(4):2141–68.
Paulk, Mark C., Bill Curtis, Mary Beth Chrissis, and Charles V Weber. 1993. "Capability Maturity Model, Version 1.1." *IEEE Software* 10(4):18–27.
Peffers, Ken, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. 2007. "A Design Science Research Methodology for Information Systems Research." *Journal of Management Information Systems* 24(3).
Robert Jr, Lionel P., Gaurav Bansal, and Christoph Lütge. 2020. "ICIS 2019 SIGHCI Workshop Panel Report: Human--Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence." *AIS Transactions on Human-Computer Interaction* 12(2):96–108.
Rothenberger, Lea, Benjamin Fabian, and Elmar Arunov. 2019. "Relevance of Ethical Guidelines for Artificial Intelligence--a Survey and Evaluation." in *ECIS 2019 Proceedings*.
Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not so Different after All: A Cross-Discipline View of Trust." *Academy of Management Review* 23(3):393–404.
Sadiq, Raghad Baker, Nurhizam Safie, Abdul Hadi Abd Rahman, and Shidrokh Goudarzi. 2021. "Artificial Intelligence Maturity Model: A Systematic Literature Review." *PeerJ Computer Science* 7:e661.
Shneiderman, Ben. 2020. "Bridging the Gap between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10(4):1–31.
Siau, Keng, and Matti Rossi. 1998. "Evaluation of Information Modeling Methods-a Review." Pp. 314–22 in *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*. Vol. 5.
Smuha, Nathalie A. 2019. "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence." *Computer Law Review International* 20(4):97–106.
Someh, Ida, Barbara Wixom, and Angela Zutavern. 2020. "Overcoming Organizational Obstacles to Artificial Intelligence Value Creation: Propositions for Research." in *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
Strauss, Anselm, and Juliet Corbin. 1994. "Grounded Theory Methodology: An Overview."
Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. 2021. "Trustworthy Artificial Intelligence." *Electronic Markets* 31(2):447–64.
Uren, Victoria, and John S. Edwards. 2023. "Technology Readiness and the Organizational Journey towards AI Adoption: An Empirical Study." *International Journal of Information Management* 68:102588.
Venable, John, Jan Pries-Heje, and Richard Baskerville. 2016. "FEDS: A Framework for Evaluation in Design Science Research." *European Journal of Information Systems* 25(1):77–89.
Wamba-Taguimdje, Serge-Lopez, Samuel Fosso Wamba, Jean Robert Kala Kamdjoug, and Chris Emmanuel Tchatchouang Wanko. 2020. "Influence of Artificial Intelligence (AI) on Firm Performance: The Business Value of AI-Based Transformation Projects." *Business Process Management Journal* 26(7):1893–1924.
Wang, Pei. 2019. "On defining artificial intelligence." *Journal of Artificial General Intelligence* 10(2): 1-37.
Wolfswinkel, Joost F., Elfi Furtmueller, and Celeste P. M. Wilderom. 2013. "Using Grounded Theory as a Method for Rigorously Reviewing Literature." *European Journal of Information Systems* 22(1):45–55.
Yin, Robert K. 2003. "Designing Case Studies." *Qualitative Research Methods* 5(14):359–86.
Yu, Xinying, Shi Xu, and Mark Ashton. 2023. "Antecedents and Outcomes of Artificial Intelligence Adoption and Application in the Workplace: The Socio-Technical System Theory Perspective." *Information Technology & People* 36(1):454–74.
Zhu, Liming, Xu, Xiwei, Lu, Qinghua, Governatori, Guido, and Whittle, Jon. 2022. "AI and ethics—Operationalizing responsible AI." *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership*, 15-33.