Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

Developers’ and accounting practitioners' perceptions of trust in AI as intelligent automation

Relevance of the chosen topic

Novel AI systems and applications are anticipated to cause fundamental changes in work-life, either automating different types of work tasks or augmenting human capabilities (Brynjolfsson & McAfee, 2011, 2017; Jarrahi, 2018). This technological change has been defined as “fourth industrial revolution” and it is predicted to result, for instance, as reskilling and reskilling of domain experts. However, this change is not sudden or happen in a void. First the domain experts must be capable and willing to use novel AI applications. This change has raised the discussion of trust as trust is considered essential for the use and acceptance of new technology (McKnight et al. 2011; de Visser et al. 2018; Siau & Wang 2018; Gefen et al., 2003; Pavlou & Gefen, 2004). But how should we perceive the concept of trust in autonomous, adaptive, and possibly unpredictable AI systems? Previous research on trust in technology does not provide answers to these questions focusing mainly on the techno-centric perspective and therefore, excluding contextual and social elements on technology trust formation.

In this qualitative case study, we examine domain expert users’ trust in novel intelligent automation in the professional context of accounting. This AI application aims to automate purchase invoice processing. By tracking accountant practitioners’ work AI application creates company-specific rules. These rules are ensured and qualified by the accountant practitioners during invoice processing, and any changes they make, are modified into these company-specific rules. As an outcome, this interaction is supposed to lead to intelligent automation, where purchase invoices can be processed fully automatically. This AI application offers an interesting case to study trust because it includes novel AI

Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

characteristics: learning from the user and removing a certain work task. AI is not ready to use when implemented into organizational practices which differs from traditional, non-intelligent technology that has typically been used to study trust in technology.

Using semi-structured interviews, we explore accounting practitioners' (hencefort, accountants) experiences and perceptions of trust in intelligent automation in their work. In addition, we explore the AI developers' aims to increase their clients' trust in AI, and how these perceptions meet during the deployment process. The preliminary results emphasize trust in technology as socially shaped and highly influenced by the specific use context and domain expertise. For instance, AI developers emphasize resistance and social dynamics that prevent trust in AI whereas accountant practitioners underline personal experience of the AI system's usefulness in their work. The findings underline fundamental differences in the accountants' and the developers perceptions of trust in AI in accounting.

This study contributes to human-centered research of AI design and development. It is noteworthy to critically consider the role of trust in the constantly digitalized work environment and the balance between overtrust and distrust in technology. Trust happens in an interaction with the "other" and the trusted party can be either another human or an artificial agent. We believe that the dynamic nature of trust and the sociotechnical essence of novel AI applications set the need to study trust in technology from an empirical and a sociotechnical perspective. This study concretizes the topic of trust in intelligent automation focusing on a specific user group and context, shedding light on this on-going phenomenon and increasing understanding of human-AI interaction and collaboration.

Research question

Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

How do accounting practitioners and developers perceive trust in AI as intelligent automation, and how do these perceptions encounter each other?

Theoretical background

Trust can be approached from many perspectives and it can be studied both qualitative and quantitatively. Despite the perspective, there are certain elements that define a trust relationships. Firstly, trust is interactive as it is formatted in a relationship with the "other", for instance, another person or artificial agent. Secondly, trust is uncertain by nature, and entering into trust relationship includes risks if trust is violated. Thirdly, trust is goal-oriented: a trust relationship aims to achieve a particular action or outcome.

The theoretical background of this study builds on Information Science, Human-Computer Interaction and Human-centered AI. Previous research on trust in technology has adapted a definition of trust from organization and management science where trust is defined 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 party."* (Mayer et al. 1995). Generally, trust is recognized as a key element impacting the user uptake and acceptance of any new technology (McKnight et al. 2011; de Visser et al. 2018; Siau & Wang 2018; Gefen et al., 2003; Pavlou & Gefen, 2004).

Previous research on trust in automation offers a useful background to study trust in autonomous and adaptive AI technology. According to Lee & See (2004), two critical elements define the basis of trust. The first is the focus of trust: what is to be trusted? The second is the type of information that describes the entity to be trusted: what is the information supporting trust? This information guides expectations regarding how well the

entity can achieve the trustor's goals. Additionally, previous research emphasizes the dimensions of performance, process, and purpose on trust in automation (Lee and Moray, 1992; Lee and See, 2004; and Madhavan and Wiegmann, 2007). Performance refers to the current and historical operation of the automation and describes what the automation does. Process is the degree to which the algorithms of the automation are appropriate for the situation and able to achieve the operator's goals, describing how the automation operates. Purpose refers to the degree to which the automation is being used according to its' intent, describing why the automation was developed.

The performance-process-purpose framework has also been used to study trust in AI, for instance, by Hengstler et al. (2016). Using a qualitative case study approach, they explored how firms systematically foster trust regarding applied AI. According to their findings, trust in applied AI requires not only trust in the technology but also trust in the innovating firm and its communication. The performance basis is primarily reliant on both operational and data security aspects, the process basis is determined by cognitive compatibility, trialability, and usability, and the purpose basis is founded on application context and design. Trust in the innovating firm increases with stakeholder alignment, high public transparency of the development project, and gradual introduction of the technology. Ultimately, trust in communication grows primarily by early, proactive, and application-based communication as well as the transmission of benefit-related information. The findings of this study suggest that trust in AI is not based only on technical characteristics but is formatted also on interpersonal communication and collaboration.

It is noteworthy that the previous research on trust in technology has mainly focused on technical characteristics, excluding contextual and empirical perspectives on trust formation. Few studies have made attempts to study trust in AI from empirical perspective

Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

(e.g., Araujo et al. 2020, Andras et al. 2020, Jacobs et al. 2021). For instance, Jacobs et al. (2020) investigate the contextual factors in clinicians' trust in decision-support tools. Their study was motivated by the lack of context awareness of the broader sociotechnical systems in which these tools are being embedded. We argue that the novel AIs' unpredictable and societal nature set the need to re-investigate trust in technology from a sociotechnical and human-centered perspective. Human-centered AI is a novel approach that underlines reliable, safe and trustworthy AI systems with an aim to amplify, augment, empower, and enhance human performance (Shneiderman, 2020).

Methodology

This study is a qualitative case study, including semi-structured expert interviews and a 3-hour workshop with the developers from the developing company. The interviewed domain expers (n=9) are accountants who are using this specific AI application in their work. These interviews focused on trust formation in AI as intelligent automation, work practice, and the overall deployment process. Company workshop was organized before conducting the expert interviews to understand the developers' perception of trust formation and the functionality and purpose of the AI application.

The data is descriptive, and the preliminary analysis is conducted with Atlas.ti using interpretive thematic analysis methods. Thematic analysis is used to identify and analyze patterns of meanings and experiences (themes) regarding the processes and related attitudes (Adams et al., 2018). Transcripted data is carefully read and re-read followed by descriptive and sentence-level open coding. Initial codes are summarized into categories, in which the emerging themes are identified. Themes are developed further and reviewed against the coded data and the overall data set, followed by a detailed analysis and definition of the

Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

themes. Finally, the analysis is contextualized in relation to existing literature and reported in a research article. The analysis aims to understand trust in AI as intelligent automation both from the end-users and the developers' perspectives, and to explore how these perceptions meet. Analyzing these perceptions helps to understand trust formation in this sociotechnical context and included interactions.

Main findings

This study aims to gain an overall understanding of trust formation during the implementation process, and the social dynamics influencing the use and acceptance of the AI as intelligent automation in this socio-technical environment. Findings are based on the preliminary thematic analysis and introduce the accountants' perceptions of trust, including trust in technology and trust in developers, and the developers' perceptions of their clients' trust formation. Before presenting the main findings, the overall deployment process and the purpose of the AI application is shortly explained.

Describing the implementation process

The AI application is introduced to the end-users (accountants) after the decision of the AI deployment has already been agreed with client company managers. In the first meetign with the end-users (acocuntants), the developers introduce the application and its functionalities. This meeting focuses on practicalities: how the AI works, what kind of clients are suitable to automate with the AI, and what clients cannot be automatized with the AI. After the application has been used for ca. 2 months, a second meeting with the accountants is organized. This meeting focuses on fine adjustments: developers aim to encourage accountants to use the intelligent automation and help them to overcome possible challenges

Vipuvoimaa EU:lta 2014–2020
Euroopan unioni
Euroopan aluekehitysrahasto

or issues during the use. This is followed by regular monitoring of the use and a push for increasing automation. The deployment process is very customer-centric and the application is tailored to support existing organizational processes and practices: invoice automation can be integrated into an existing system or deployed in a separate interface that is developed around the AI application.

The implementation of the AI application is gradual and includes three phases. The process starts from (1) automated learning. The AI tracks the accountant's work and learns on a customer-specific basis how accounting, cost-centers, routing, and VAT processing have been done in the past and automatically creates company-specific rules. This is followed by (2) quality assurance. The AI application starts processing invoices and the accountant ensures the rules work. If the accountant makes any changes, the developing company modifies the company-specific rules based on the changes made by the accountant. The third step of the process is (3) intelligent automation. Those invoices that do not need any input from the accountant will be processed fully automatically by the developing company and will not be shown to the accountant.

Accountants' perceptions of trust

At the beginning of the deployment process, the accountants' trust in technology developers emphasize. Before trusting the AI, the participants evaluate the developers' competence to design and develop the AI application: do they understand the context thoroughly and are capable of developing efficient appplications for accountants' work. The accountants emphasize participatory and collaborative design processes and underline the importance share their knowledge and understanding about the use context and the domain with the developers. In addition, the accountants appreciate the opportunity to influence the

application development, for instance, by sharing their expertise with the AI developers. They feel important that technical support and help are available when needed, and that the developers are honest and open regarding the AI applications' cababilities and potential pitfalls.

Trust in AI, in human-AI interaction, focuses mainly on the technical functionality of the system. The AI application is considered trustworthy if it operates as expected, meets the user expectations and does result in extra work. Trust formation in AI is gradual. Firstly, accountants need to have a thorough understanding of technology, its abilities and technical functionality. Secondly, accountants need to have a thorough client-specific understanding and knowledge. The main elements of this assurance are invoice stability and repetition. Lastly, based on this information, they can decide if their clients' invoice processing can be automatised or not, and if yes, in which scale. However, this is just a starting point for trust formation. Accountants indicate a careful attitude when implementing automation into their work practices and underline a need to evaluate and double-check the applications' operation and performance. Evaluation emphasizes at the beginning of the implementation process and is expected to diminish over time as the accountants are assured of the AI applications' correct operation. It is noteworthy, that this assurance is very personal because each accountants have different clients and their use of the AI and intelligent automation differ from each other. The user interface is designed to support this assurance by indicating a percentage accuracy of the automation. However, this system property is not fully trusted, as one participant indicates: *"even though it is 100%, I trust it 90%."* The accountants underline that double-checking the accounts does not really consume a lot of time but saves their time regarding upcoming work tasks if the AI application would be proven erroneous later.

The cautious attitude might be explained by the AI applications' merging with the accountants' expertise: accountants' considered themselves responsible for the AI outcomes, and therefore need to have the assurance of the applications' accuracy. Trust in technology is seen to reflect their trustworthiness as domain experts. This is an interesting observation as it reflects the synergy that results from human-AI interaction. To accept the AI application and collaborate with it, the domain expert must first have the user-specific knowledge and the assurance about the AI applications' suitability to their work practices: the decision of how to use the AI and with which clients is very personal, even though this was discussed also in the first meeting with the developers.

Developers' perceptions of trust in AI

The developers underline the importance to educate the domain experts to use the AI application in the correct way and to achieve the full potential of automation capabilities. However, education does not focus only on the use of AI application. The developers underline a need for change management, as a wider purpose of the AI is to modernize and digitize accountants' work through intelligent automation. The developers feel the need to push the accountants to change their work practices and demonstrate resistance among accountant's toward this change. According to developers, domain expert users need so-called "leap of faith" to understand the benefits of automation and to trial the AI application but possible prejudices might prevent the use of automation. These prejudices might be associated with the lack of identifying potential benefits the automation, or the fear of losing their work to automation. Gaining users' trust in the early phase of the AI deployment is seen as crucial in helpful to overcome these challenges in the use and acceptance of the AI.

As indicated, the accountants can decide the scale regarding the invoice automatization, and the automation can be increased or decreased according to accountants'

individual preferences. In addition, the client organizations' managers also monitor the invoice processing: if they identify opportunities to increase automation, they can encourage accoutants to act on it. However, the developers are planning to change this feature from opt-in to opt-out: they aim to push the use of automation during a fixed time period, and the accountants should opt out if automation is not possible with some of their clients. This is seen as a solution to increase the use of automation: the current process is not seen as effective enough, although the developers also underline the importance to give the end-user the option of declining automation. In addition, social dynamics inside an organization are considered to influence trust formation in AI and intelligent automation. For instance, persons with negative attitudes and distrust towards AI application inside an organization might influence also other users perceptions of the too. To overcome this, the developers have utilized some of the end-users to work as "ambassadors" who could encourage and help their colleagues to overcome possible challenges when using the AI application.

Summary

The findings of this study indicate that trust in technology, and especially trust in AI and intelligent automation is not only technical but also social and interpersonal. Trust perceptions seems to counteract between two perspectives: The developers underline a deterministic perspective where technological change is seen as inevitable. They emphasize the change management and users' resistance, attitudinal aspects and underlying social dynamics either supporting or preventing the deployment process and trust formation. Willingness to take a "leap of faith" to trial the AI application is seen crucial in this change. Accountants, on the other hand, underline the social constructivist approach. They evaluate the developers' capability and competence to develop and design AI application to the specific use context, and their ability to collaborate in, and influence the design process and

the applications' properties. They emphasize the need for a thorough understanding of their work practices and client-specific information, before implementing intelligent automation to their work. Trust formation is gradual, and is increases as the AI applications' functional operation is personally evaluated and guaranteed. Balancing these two fundamentally different approaches require empathy, understanding, and a human-centered perspective in AI design and development – and more research on sociotechnical trust formation.

Contribution

This study aims to concretize trust formation in AI and intelligent automation in specific use context and between different stakeholders. The findings indicate that trust in technology is not built solely on human-AI interaction but also in interpersonal communication between other stakeholders in this AI-ecosystem, such as in the interaction between domain-expert users and technology developers. Also, the perception of trust differs between these two groups. This case study aims to understand the sociotechnical elements in trust in technology focusing on a wider context instead of technical features, expanding the research on trust in technology towards more sociotechnical approach. When designing novel AI application and systems into future work, we are designing society as well. I argue that this change should not consider only technical features of trust formation but also, and maybe even more, social and societal.

References

Adams, A., Lunt, P., & Cairns, P. (2008). A qualititative approach to HCI research. In: Cairns, Paul and Cox, Anna eds. Research Methods for Human-Computer Interaction. Cambridge, UK: Cambridge University Press, pp. 138–157.

Andras, P., Esterle, L., Guckert, M., Han, T. A., Lewis, P. R., Milanovic, K., … Wells, S. (2018). Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems. IEEE Technology and Society Magazine, 37(4), 76–83. https://doi.org/10.1109/MTS.2018.2876107

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI and Society, 35(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Brynjolfsson, E., McAfee, A., (2011). Race against the machine: how the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Digital Frontier Press, Lexington, MA.

Brynjolfsson, E., & Mcafee, A. (2017). The Business of Artificial Intelligence: what it can and cannot do for your organization. Harvard Business Review Digital Articles.

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction. Ergonomics, 61(10), 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and tam in online shopping: AN integrated model. MIS Quarterly: Management Information Systems, 27(1), 51–90. https://doi.org/10.2307/30036519

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices. Technological Forecasting and Social Change, 105, 105–120. https://doi.org/10.1016/j.techfore.2015.12.014

Jacobs, M., He, J., & Pradier, M. F. (2021). Designing ai for trust and collaboration in time-constrained medical decisions: A sociotechnical lens. In Conference on Human Factors in Computing Systems - Proceedings. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445385

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons, 61(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. In Ergonomics (Vol. 35, Issue 10, pp. 1243–1270). https://doi.org/10.1080/00140139208967392

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277–301. https://doi.org/10.1080/14639220500337708

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20, 709–734.

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. ACM Transactions on Management Information Systems, 2(2). https://doi.org/10.1145/1985347.1985353

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. Information Systems Research, 15(1), 37–59. https://doi.org/10.1287/isre.1040.0015

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human-Computer Interaction, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal, 31(2), 47–53.