Amongst a Multitude of Algorithms: How Distrust Transfers Between Social and Technical Trust Referents in the AI-Driven Organization

Rebecka C. Ångström
Stockholm School of Economics
rebecka.cederingangstrom@phdstudent.hhs.se

Magnus Mähring
Stockholm School of Economics
magnus.mahring@hhs.se

Martin W. Wallin
Chalmers University of Technology & ETH Zurich
martin.wallin@chalmers.se

Eivor Oborn
Warwick Business School
eivor.oborn@wbs.ac.uk

Michael Barrett
Cambridge Judge Business School
m.barrett@jbs.cam.ac.uk

Abstract

Although trust is identified as critical for successfully integrating Artificial Intelligence (AI) into organizations, we know little about trust in AI within the organizational context and even less about distrust in AI. Drawing from a longitudinal case study, in which we follow a data analytics team within an organization striving to become AI-driven, this paper reveals how distrust in AI unfolds in an organizational setting shaped by several distrust dynamics. We present three significant insights. First, distrust in AI is situated and involves both social and technical trust referents. Second, distrust is misattributed when a trust referent is rendered partly invisible to the trustor. Finally, distrust can be transferred between social and technical trust referents. We contribute to the growing literature on integrating AI in organizations by presenting a model of distrust transference activated by social and technical trust referents.

Keywords: Artificial intelligence, trust, distrust transference, social and technical trust referents, AI-driven organizations.

1. Introduction

As organizations launch initiatives to become AI-driven (Agrawal et al., 2018; Iansiti & Lakhani, 2020), they commonly introduce artificial intelligence (AI) to automate and transform work (Berente et al., 2021; Rai et al., 2019; von Krogh, 2018). Alongside these developments, a spectrum of concerns has come to the fore regarding the consequences of using AI technologies in and for organizations, including how AI may influence job content, job security, and human autonomy (Christin, 2017; Kellogg et al., 2020). It is, thus, not surprising that practitioners and scholars alike have pointed out the importance of trust for the successful integration of AI into the workplace (Fountaine et al., 2019; Glikson & Woolley, 2020; Leonardi et al., 2022), and failure to establish trust in AI contributing to rejection or disuse of the technology (Brayne & Christin, 2021; Dietvorst et al., 2015).

With few exceptions (Leonardi et al., 2022; Lumineau et al., 2022; Söllner et al., 2016), however, studies on AI and trust have focused on the direct relationship between an individual human trustor and a specific AI artifact (Jacovi et al., 2021; Lockey et al., 2021), rather than on how trust is shaped in an organizational context. This oversight is unfortunate since developing AI often engages individuals and units across organizational domains (Fountaine et al., 2019; Iansiti & Lakhani, 2020).

Furthermore, although the absence of trust has been identified as driving the rejection of AI (Dietvorst et al., 2015), the role of active distrust of AI has remained relatively unexplored. From the trust literature (Lewicki et al., 1998; Mayer et al., 1995; Rousseau et al., 1998), we know that disruptive events, such as organizational transformation and technological advancement, can threaten employee trust in the organization (Dirks & de Jong, 2022; Gustafsson et al., 2021; Kähkönen et al., 2021), and has led to distrust between groups (Sørensen et al., 2011).

Thus, in this paper, we take a more holistic approach to explore how distrust in AI is shaped when organizations undertake efforts to become AI-driven. Our research question was: How do distrust dynamics unfold while integrating AI tools and AI-related work practices into the organization?

We conducted a longitudinal case study at a business unit within a multinational technology firm (Global Tech) undergoing a significant transformation to become AI-driven. We followed the work of a data

URI: https://hdl.handle.net/10125/107079
978-0-9981331-7-1

analytics team with the assignment to develop an extensive range of algorithms, serving on the frontline for realizing a corporate AI initiative, including their interactions with users.

During our fieldwork, we identified distrust phenomena related to AI development that remained unresolved despite the developers' best efforts. We noticed that neither the literature on trust in AI nor the organizational trust literature seemed to adequately explain these observations, leading us to focus our investigation on distrust in relation to AI.

Our findings reveal that distrust in AI was situated and involved both social and technical trust referents. We showed that when a trust referent is rendered partly invisible to the trustor, it leads to misattribution of distrust. We also showed that distrust can be transferred between social and technical trust referents. Based on these findings, we make two key contributions. First, we contribute to the growing literature on integrating AI in organizations (Berente et al., 2021; Faraj et al., 2018; van den Broek et al., 2021) by articulating a richer understanding of the crucial role of distrust in AI and developing a model of distrust transference actuated by partly invisible social and technical trust referents. Second, we contribute to the literature on trust (Fulmer & Gelfand, 2012; Lumineau et al., 2022) by articulating how digital artifacts are integral to shaping organizational trust relations.

2. Literature

2.1 What is AI?

Following Faraj et al. (2018), we use the term AI to refer to "an emergent family of technologies that build on machine learning, computation, and statistical techniques, as well as rely on large data sets to generate responses, classifications, or dynamic predictions that resemble those of a knowledge worker" (Faraj et al., 2018, 62). AI is different from traditional information technologies in its ability to digest vast amounts of data to identify patterns, predict outcomes, and propose proactive solutions (Agrawal et al., 2018; Faraj et al., 2022).

Through data, AI can continue to learn and improve its accuracy (Berente et al., 2021). The dependency on data is also AI's vulnerability. Learning from low-quality data containing faults, biases, or missing data points will deteriorate AI reliability and potentially lead to algorithmic breakdowns (Boyd & Crawford, 2012; Danks & London, 2017; Faraj et al., 2018). We use the term algorithm when referring to a specific AI application.

2.2 Trust in AI

As a general-purpose technology, AI is expected to be applied in various fields within the organization, automating or aiding cognitive tasks, such as decision-making (Agrawal et al., 2018; Brynjolfsson & Mitchell, 2017; Rai et al., 2019; von Krogh, 2018). Integration can also result in adverse effects, such as job loss (Frey & Osborne, 2017) or increasing surveillance and control over employees (Brayne & Christin, 2021; Faraj et al., 2022; Kellogg et al., 2020). With AI expected to impact organizations broadly, scholars and practitioners have pointed out the importance of trust for the successful integration of AI (Fountaine et al., 2019; Glikson & Woolley, 2020; Leonardi et al., 2022).

Trust is commonly defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau et al., 1998, 395). Trust involves a trustor, the party that is trusting, and a trust referent or trustee, the party that is trusted. The perceived trustworthiness of a trust referent depends on factors such as the trust referent's ability, benevolence, and integrity (Mayer et al., 1995).

Regarding AI, the trust relationship usually includes an individual human trustor and a single algorithm as the trust referent (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021). The perceived trustworthiness of an AI is dependent on system-like trust constructs, such as the AI functionality, reliability, and helpfulness (Lankton et al., 2015), but also its transparency and the level of task substitution (e.g., automation or augmentation) (Glikson & Woolley, 2020; Lockey et al., 2021). Violating these trust constructs can give rise to negative expectations or rejection of an algorithm, known as algorithmic aversion (Dietvorst et al., 2015; Glikson & Woolley, 2020).

2.3 Distrust in AI

Distrust is defined as "confident negative expectation regarding another's conduct" (Lewicki et al., 1998, 439) where the trustor is unwilling to succumb to vulnerability (Bijlsma-Frankema et al., 2015; Lewicki et al., 1998). Distrust is not equivalent to low trust. Instead, it is a separate construct from trust (Dimoka, 2010; Lewicki et al., 1998; Saunders et al., 2014), following different dynamics and potentially occurring simultaneously (Komiak & Benbasat, 2008). For example, studying the usage of a recommendation agent (RA), Komiak & Benbasat (2008) found that distrust was built up when users became aware of information unknown to them, when perceiving the RA as incompetent, and when the RA did not meet their expectations; trust was developed when the users perceived the RA as competent and providing sound and adequate information. Similarly, distrust can emerge during disruptive periods, such as organizational transformations and technological advancements, where distrust can develop in self-amplifying cycles (Bijlsma-Frankema et al., 2015). For instance, uncertainty during an organizational change program can lead employees to interpret management's intentions negatively, and management can interpret employee reactions negatively, feeding a cycle of trust deterioration and distrust development (Sørensen et al., 2011).

Within organizations, employees can reject algorithms and develop strategies to resist their influence (Christin, 2017; Kellogg et al., 2020). Domain experts have been found to be more reluctant than laypeople towards algorithms augmenting or automating tasks (Hoff & Bashir, 2015; Logg et al., 2019). A reason for domain experts' skepticism can be that they perceive a risk of deskilling and loss of job security as algorithms start to perform tasks independently and, as such, will compete with domain experts (Lockey et al., 2021).

2.4 Social relations influence on trust and distrust in AI

The question of how additional trust relations, beyond the individual trustor and the AI trust referent, influence trust in AI is relatively unexplored. Recent studies, however, have begun to recognize AI developers as relevant to trust in AI (Hengstler et al., 2016; Leonardi et al., 2022; Lumineau et al., 2022, Söllner et al., 2016), while portraying them as anonymous and distant (Lumineau et al., 2022; Söllner et al., 2016).

However, the organizational trust literature tells us that trust relations can form intricate webs involving trustors and trust referents across different analysis levels (Fulmer & Gelfand, 2012; Lumineau & Schilke, 2018). Since the situated development of AI engages both technical and domain experts, we argue that it becomes important to further investigate how social trust relations within the organization influence trust and distrust in AI.

2.5 Trust transference

Trust transference is a cognitive process where "the trustor transfers trust from a known entity to an unknown one" (Doney et al., 1998, 605), or, put differently, when a trustor bases their initial trust in one party (individual, team, or organization) on their trust in another party (Stewart, 2003). Trust transference can occur when there are differences in the level of trust between trust referents, such as during trust repair (Bachmann, 2015; Kähkönen et al., 2021). Similarly, distrust can transfer from a distrusted party, damaging the legitimacy of a credible party (Bachmann et al., 2015).

For (dis)trust transference to occur, the trustor must be able to establish links between the parties in question (Doney et al., 1998), perceiving them to be related, for instance, by their similarity, proximity, or common view or interests (McEvily et al., 2003; Stewart, 2003). Research has also explored trust transference in relation to trust referents such as technology, providers, and platforms, as well as users' trust in specific services, like public e-services (Belanche et al., 2014), mobile payments (Gong et al., 2020), platforms (Chen et al., 2015; Shao et al., 2022) and self-driving vehicles (Renner et al., 2022).

3. Method

3.1 Empirical context

Our research site was located at a local operation center in Europe, part of a multinational technology company ("GlobalTech"). The operation center managed geographically dispersed installations of field equipment for GlobalTech's customers. This work included supervising the equipment, responding to equipment alarms, and dispatching and supporting field technicians serving the equipment. During our fieldwork, the operation center was transformed significantly to become AI-driven. The corporate AI strategy was manifested in a new operating model, which included transforming the organizational structure, assignments, and roles.

Numerous teams were involved in the operational work serving customer equipment, from operative teams monitoring equipment to domain experts handling critical incidents. We refer to all these teams as 'operations teams' unless it is relevant to point out their specific functions. Situated within the local organization was a data analytics team ("DA team"), the only team with technical expertise in data science. The DA team was assigned to deliver various data analytics models ("algorithms") to support the operations teams, using predictive modeling, advanced data analytics, and data visualization dashboards. We followed the organizational transformation to implement the corporate AI strategy, specifically narrowing in on the occurrences between the operations teams, the DA team, and the algorithms.

Furthermore, to build these algorithms, the DA team accessed data from GlobalTech systems, the customers' systems, and third parties. These data included information on equipment, alarms, work orders, and external information such as weather forecasts. The DA team was mostly externally recruited, thus lacking operations domain expertise.

3.2 Data collection

We were granted access to GlobalTech's operations center from May 2019 to December 2021, with this paper primarily focusing on events occurring between May 2019 and May 2020. During fieldwork, we collected data both onsite and online. We spent 16 days at the operations center, of which 12 for observations. We conducted 51 semi-structured interviews, 18 recorded follow-up interviews, and nine discussions documented in field notes.

The fieldwork included conversing with 32 informants, and each recorded conversation (interview or discussion) spanned between 30–120 minutes. In the semi-structured interviews, we asked our informants to reflect on previous and ongoing critical events, as well as ongoing collaboration and relations between the teams. The total recorded material is 60 hours, all transcribed. In parallel, 110 documents were collected, including reports, presentations, emails, internal news postings, and Yammer conversations. In the final edits, some of the quotes in the paper have been grammatically corrected for readability.

3.3 Data analysis

We adopted the principle of constant comparisons from grounded theory method (Glaser & Strauss, 1967), where data collection and coding are conducted iteratively as the fieldwork progress. In addition, we followed established guidelines for inductive concept development (Gioia et al., 2013). The coding progressed during the fieldwork and resulted in 122 codes. These codes included perceptions and actions such as 'Fear of, or resistance against, change,' 'The challenge of data (quality, access, structure),' 'Building trust/confidence,' and 'Being blamed.' As critical events and specific algorithms emerged from the data, we started to code them separately, allowing us to follow these items as they appeared in interviews and observations in the continuing fieldwork. This coding also allowed us to uncover temporal aspects of trust development, for instance, how specific algorithms gained trust over time. As distrust emerged as a central phenomenon, we conducted a second round of coding where we identified statements and behaviors expressing positive or negative trust perceptions (Brattström et al., 2019; Lewicki et al., 1998). All trust statements and behaviors were thoroughly examined to determine the trustor, trust referent, and relevant factor of trustworthiness (Glikson and Woolley, 202; Mayer et al., 1995). We identified four trust referents: the corporate AI strategy, the DA team, the algorithms, and the data.

Building on our initial coding, we noticed that distrust emerged as a phenomenon. Comparing with existing concepts from the literature, such as trust in AI (Glikson & Woolley, 2020), distrust development (Komiak & Benbasat, 2008; Lewicki et al., 1998; Sørensen et al., 2011), and trust transfer (e.g., Stewart, 2003), we looked for similarities and differences that could explain our phenomenon. The comparison resulted in second-order themes, distilled into three aggregated dimensions (Gioia et al., 2013), revealing three distrust dynamics that actively influence distrust during the continuous integration of algorithms. Our findings are presented thematically according to these dynamics.

4. Findings

4.1. Distrust dynamic 1: Distrust in the Corporate AI strategy needs to be handled by the DA team

On November 14th, 2018, the business unit at GlobalTech announced that they have a new corporate AI strategy that included a new operating model focusing on AI, automation, and data. It proclaimed that the operations must be rebuilt from the ground up using AI and automation as core elements. The aim was to harness AI capabilities, remove monotonous and repetitive tasks from the current working method, and improve efficiency and reduce costs. The vision of what the corporate AI strategy would bring was far-reaching, and external marketing material emphasizes futuristic AI capabilities.

For instance, in a promotion video, the head of the GlobalTech business unit was seen conversing with a futuristic AI that managed a customer's field equipment over a large geographical area. The video had the purpose to illustrate the future operational work, where the AI, equipped with a natural female voice, performed the work currently assigned to the employees at the local operating center. The only task the Head of the Business unit needed to do was verbally accept the AI's suggested actions.

The management identified that the organization pyramid was 'too fat' at the bottom, meaning there were too many low-skilled roles. They were convinced that employees should either up- or re-skill to be relevant for the new strategy. However, as the management was still determining what jobs would be available after the shift, they kept information regarding organizational structuring, roles, and tasks described at a very general level. Instead, the internal communication targeting the employees focused on the organization's reasons for change, new future tools and processes, and expectations on employees to adopt new mindsets.

The future vision of AI together with the need for more clarity regarding roles and tasks raises uncertainty amongst employees regarding the corporate AI strategy's impact on job security. The Top Manager 1 reflects on the reaction he received during the strategy rollout:

> *People recognized that no matter how well we dressed it up, their job was under threat as it stood at the time* (Top Manager 1).

The DA team welcomed the corporate AI strategy and operating model. From the start, they recognized themselves as part of the corporate AI strategy and took pride in their roles, processes, and algorithms aligning with the new operating model. At the same time, they recognized that other teams at the operation center were more reluctant toward corporate AI strategy. The DA Manager 1 expressed this connection:

> *What is data-driven? What is a proactive approach? What is automation? People are scared about that. I think it is just because there is no clear understanding* (DA Manager 1).

Seeing themselves and their algorithms as dependent on operations teams' accepting the corporate AI strategy, the DA team has arranged workshops to demonstrate predictive algorithms, explain the new corporate AI strategy's value and benefits, and promote the team and their algorithms.

To summarize this distrust dynamic: The corporate AI strategy stressed the future functionality of AI and overplayed AI capabilities. Simultaneously, practical implications are hidden, inciting uncertainty and vulnerability among the employees, who start to fear for their jobs and distrust the corporate AI strategy. The DA team actively associates itself with the corporate AI strategy. This association, however, enables distrust transfer between the corporate AI strategy and the DA team. Therefore, to successfully build trust in their algorithms, the DA team must overcome the distrust for the corporate AI strategy.

4.2. Distrust dynamic 2: Perception of the DA team leads to the rejection of algorithms

The operations teams found that the DA team were too distant from the operations to capture their needs when developing algorithms. They complained that the DA team needed to understand the operations and question whether the DA team was even interested in learning about the domain. For example, a customer officer complained that the DA team was not interested in what the operations teams—with domain expertise—wanted them to address:

> *They are just doing things on their own, without understanding what they are doing […] they are just developing something. They think that it is the way forward, and they are not listening to the, you know, the real experts.* (Customer officer)

At the same time, the operations teams are unaware of the DA teams' technical work and expertise. Most of the DA team's work included different forms of data and algorithm preparations, which are invisible to the operations teams. The DA Manager 2 explained that he believed only a small part of the team was visible to the operations team:

> *There are people doing data engineering, data modeling, and data understanding. [The operations teams] are seeing only one person, the data scientist* (DA Manager 2).

The perceived lack of domain expertise led to the operations teams' rejecting support from the DA team. This rejection became visible in a particular case. A data scientist from the DA team developed an algorithm that predicted when a piece of field equipment was at risk of malfunctioning due to hot weather. The possibility of predicting overheating equipment would have allowed the operational team to take preventive actions, for instance, sending field technicians to cool down the equipment.

However, the operational team managing the customer account was not interested. To convince them of the algorithm's value, the data scientist emailed the team every time the algorithm predicted that a piece of equipment was at risk of overheating. She sent these emails for almost a year before the team finally accepted the algorithm. Operations Manager 1 explained that the operations team resisted the model since they did not trust the data scientist to understand their needs:

> *Interviewer: Was there a lack of trust in the [algorithm] or the data scientists?*
>
> *Operations Manager 1: I think it was a lack of trust for the data scientist.*

To summarize this distrust dynamic: During collaborations, the operations team perceived that the DA team lacked domain expertise resulting in the impression that they lacked the ability to build valuable algorithms. At the same time, the operations team did not understand the work that the DA team performed or the constraints that bound it. As such, the DA team's technical expertise was invisible, hiding their actual ability. Perceiving that the DA team lacked ability resulted in operations teams distrusting the DA team. The distrust was transferred to the DA team's algorithms, manifesting as operations teams rejecting the DA team's algorithms.

4.3. Distrust dynamic 3: Data issues create distrust in algorithms and are blamed on the DA team

While managing the field equipment, operational employees interacted with different IT tools to fill in information and respond to system output. The information was used for tracking incidents, analyzing the root causes of the incidents, and communicating with other teams and customers. The employees' activities with the IT tools also generated data that the DA team extracted for their analysis and algorithms. If the operations employee was not following the operational processes, not adding standardized information, or missing adding the information altogether, it impacted the quality of the data gathered from the tools. The decrease in data quality impacted the DA team's algorithms' reliability.

For example, during the summer of 2019, one of the DA team's predictive algorithms inaccurately overestimated how long a piece of field equipment could manage without service, resulting in an equipment failure. At first, the operations team believed that an operations employee had made a mistake, but as they investigated the issue, it became clear that the algorithm's prediction was wrong. The DA team, however, conducted a subsequent investigation revealing that the algorithm had learned from data containing an operations employee mistake which was included in the algorithms training data. The DA Manager 1 commented on the human error leading to the faulty prediction:

> *We need a little bit of discipline in our work. We need to understand that if we are not disciplined with what we are doing, and we don't believe in our data, we cannot become data-driven"* (DA Manager 1).

Likewise, the operations teams' activities can impact data access. In January 2020, one of GlobalTech's call centers performed poorly, and the DA team was requested to analyze the cause. The DA team needed data from the customer's system to provide an analysis. However, accessing customer data must be handled by the operations teams, and this was not a prioritized task for them. Hence, instead of gaining access to the system directly, the DA team received different data dumps in Excel files sent daily in emails from the customer.

The lack of control and the low consistency between the files hindered the DA team from making a coherent analysis of the call center's performance. As a result, the DA teams reported it as deficient. The DA Manager 1 reflected on that the scarcity of data was the source of the rejection:

> *We are receiving by email some snapshots [Excel sheets of data]. We are running our analysis based on those snapshots. The fact that those snapshots are not complete it is not our fault. And they say, "No, the report is not good." So, there is a huge resistance* (DA Manager 1)

As algorithms were not meeting the expectation of the operations team, the operations teams started to blame the DA team. A member of the DA team reflected on how the blame shifted:

> *I mean, they are blaming the model for the problems. But the problems are not because of the model but because of the data accuracy behind the model. They are blaming the team that they didn't do a good model, but the problem, in fact, stays in the data, and this is what we tried all the time to explain; "please understand, garbage in, garbage out." Yeah, so if the data are inaccurate, don't blame the model. Yeah, don't blame the team that build the model.* (DA Employee 1)

To summarize this distrust dynamic: The operational teams influenced the algorithm's reliability and functionality by impacting data. They unknowingly did this, as their work impacted data quality and access. The algorithms' invisible dependencies led to misattributed distrust, where operations teams assigned the fault to the algorithms. Their distrust was transferred from the algorithms to the DA team as they started questioning the DA team's ability to build reliable algorithms.

5. Discussion

We began this paper with a simple observation; Prior research has argued that failing to establish trust in AI can result in rejection or disuse of the technology (Brayne & Christin, 2021; Dietvorst et al., 2015). Rallying around such consensus, a large body of research has emerged that examines AI and trust between the individual human trustor and a single algorithm trust referent (Jacovi et al., 2021; Lockey et al., 2021). However, this perspective is severely limiting for two reasons. First, the construct of distrust is seldom explored as part of the algorithmic rejections. Second, the efforts to integrate AI are often part of corporate-wide initiatives involving a multitude of algorithms and spanning individuals, teams, and units, likely creating an intricate web of relationships (Fountaine et al., 2019; Iansiti & Lakhani, 2020).

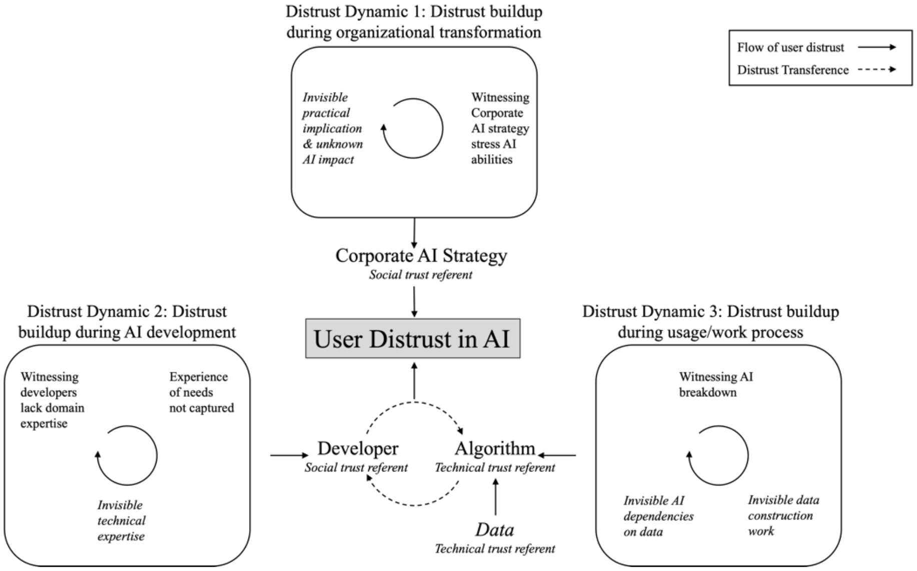
Distrust Dynamic 1: Distrust buildup during organizational transformation
Invisible practical implication & unknown AI impact
Witnessing Corporate AI strategy stress AI abilities
Flow of user distrust
Distrust Transference
Corporate AI Strategy
Social trust referent
User Distrust in AI
Distrust Dynamic 2: Distrust buildup during AI development
Witnessing developers lack domain expertise
Experience of needs not captured
Invisible technical expertise
Developer
Social trust referent
Algorithm
Technical trust referent
Data
Technical trust referent
Distrust Dynamic 3: Distrust buildup during usage/work process
Witnessing AI breakdown
Invisible AI dependencies on data
Invisible data construction work

Figure 1. A model of AI-related distrust dynamics.

To address this gap, we investigated how social and technical distrust dynamics unfold when integrating AI into organizations. We developed a range of explanations rooted in a sociotechnical (Mumford, 2006) understanding of trust in AI, incorporating social and technical trust referents and considering them interrelated. First, we identified that distrust in AI involved social and technical trust referents. Second, we recognized distrust emerged when trust referents were not completely visible to the trustor. Third, we showed that distrust was transferred between the social and technical trust referents. Based on our insights, we have presented a model displaying the three distrust dynamics that unfolded during the development and integration of multiple algorithms into the organization (see Figure 1), which we discuss below.

5.1. Distrust in AI depends on both social and technical trust referents

Our first finding places AI in a situated context, revealing how social and technical actors trigger distrust in AI that emerges, forms, and blends into the organization. We identified two social trust referents, the ‘corporate AI strategy’ and the ‘developers,’ and two technical trust referents, the ‘algorithms’ and ‘data.’

Research in organizational trust has revealed that a corporate strategy containing operational and human resources strategies could affect employees' perception of the organization's trustworthiness (Gillespie & Dietz, 2009). Furthermore, employees’ trust in organizations and managers could become challenged during major organizational transformations (Gustafsson et al., 2021; Sørensen et al., 2011).

Our research revealed a similar pattern where the organization’s decision to become AI-driven resulted in uncertainty among employees. However, we also saw that introducing AI added a new dimension of uncertainty based on the unknown potential of AI to overtake tasks and job roles. This uncertainty grew as corporate communication portrayed AI as futuristic while practical implications for job impact were invisible in the corporate AI strategy. We referred to this development as distrust dynamic 1, shown in our model. This dynamic generates distrust in the corporate AI strategy, including the organization's intention and articulated AI potential. We therefore concluded that a corporate AI strategy can (and is likely to) function as a trust referent.

Our second identified trust referent were the developers (cf. Leonardi et al., 2022; Lumineau et al., 2022). This allowed us to articulate the role of developers, situated within the organization and collaborating with surrounding teams, in AI distrust dynamics. We identified the developers as trust referents at the team level within the organization (Fulmer & Gelfand, 2012). Drawing from the trust literature, the experience of trusting a party will influence the trustor’s perception of a trust referent for future occasions (Mayer et al., 1995). This dynamic is relevant as organizations strive to become AI-driven, since relations between domain and technical experts can unfold over numerous AI development projects.

Our technical trust referents, the algorithms and the data, are related since algorithms depend on data to learn and undertake tasks (Faraj et al., 2018). We know from existing research that the algorithms’ trustworthiness depends on algorithms' capabilities (Glikson &

Woolley, 2020; Lockey et al., 2021). For instance, experiencing an algorithm having a reliability breakdown can create algorithmic aversion (Dietvorst et al., 2015; Komiak & Benbasat, 2008). However, research seldom widens the scope to explore causes behind such breakdowns (Glikson & Woolley, 2020). By separating the algorithms from the data, our research revealed that they are different trust referents and that distrust in one of them did not necessarily mean distrust in the other. We also identified that the perception of the algorithm's trustworthiness is dependent on data, which is, to our knowledge, seldom discussed in the literature on trust or distrust in AI (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021).

5.2. The partial invisibility of trust referents results in misattributed distrust

Our second insight was that when trust referents were less than fully visible to trustors, it could result in the misattribution of distrust. A particularly interesting aspect of this is the invisibility of data. Scholars have pointed out that data curation can be invisible (Sachs, 2020; Waardenburg et al., 2022) and that data curation work can be performed by invisible workers (Kellogg et al., 2020). Data construction can also be part of employees' daily work (Waardenburg et al., 2022). We expanded on this research by identifying how data and related data work can be invisible also to the people performing it.

The invisibility of data results in algorithms becoming partly invisible too, which has consequences for AI. In our research, we identified two ways in which this plays out. First, when operations employees did not see how their work impacted data quality, they did not know that they contributed to algorithmic breakdowns. Instead, they blamed the algorithms for poor reliability. Second, when they did not see how their work constrained data access, impacting algorithms, they did not challenge data scarcity. Instead, they blamed the algorithms for their lacking functionality.

The invisibility of data also affected the developers as trust referents. When operations employees noticed that the developers lacked domain expertise but failed to see the developers' technical expertise, they also failed to recognize the technical constraints that limited the development of algorithms. Instead, they blamed the developer's ability to develop algorithms.

The challenges with invisible data and data work are illustrated in distrust dynamics 2 and 3. The misattribution of distrust to algorithms and developers further supports our argument that trust and distrust in AI must be studied beyond the individual relations between a human trustor and the AI trust referent (Glikson & Woolley, 2020; Lockey et al., 2021).

5.3. Dependent trustworthiness enables distrust transfer

Distrust transference occurs, as noted, when a trustor's distrust towards one trust referent transfers to another trust referent (Bachmann et al., 2015; Doney et al., 1998), which can occur when trust referents are perceived as related, for instance, by their similarity and proximity (McEvily et al., 2003; Stewart, 2003). Our study shows that the relatedness between developers and algorithms enabled distrust to transfer between the two, as shown in our model.

Our study also shows that the relatedness between developers and the corporate AI strategy forced the developers to counteract distrust in the corporate AI strategy, avoiding the distrust to transfer to themselves and their algorithms. As such, our research reveals that distrust transfer can occur between social and technical trust referents within the organization.

Furthermore, in contrast to previous research exploring distrust cycles (Bijlsma-Frankema et al., 2015; Sørensen et al., 2011), we show that the distrust continues despite the best effort of the developers to overcome the distrust. Our model shows how distrust cycles can be fueled by distrust transference across several related trust referent. For instance, when trustors develop distrust in developers' ability, this can be transferred to the algorithms the developers produce. Likewise, when trustors perceive that algorithms have limitations to their functionality or issues with reliability, distrust can transfer to developers.

5.4. Distrust cycles stall the transformation process

Our insights differ from those of Glikson and Woolley (2020), who found that initial trust in embedded algorithms is high but slowly deteriorates over time. Rather, similar to Christin (2017) and Kellogg et al. (2020), we found that algorithms are met with employee resistance. We expanded on these insights by connecting resistance with distrust, influenced by social and technical trust referents. We identified the source for distrust as fear of job security and partial invisibility of trust referents, which challenge the work status quo. Furthermore, we revealed that distrust transfers enable distrust cycles to occur, which forces the trust trajectory to stay low over time. This distrust cycle also stalls the organizational transformation process as the uptake of algorithms is slowed.

6. Conclusion

According to current predictions, the integration of AI in organizations will be far-reaching, with multiple algorithms employed in all parts of the organization and for various purposes (Agrawal et al., 2018; Berente et al., 2021; Iansiti & Lakhani, 2020). Furthermore, AI's unique capabilities to mimic human intelligence, aiding us in decisions making (Agrawal et al., 2018; von Krogh, 2018), will allow the technology to become woven into the social fabric of organizations.

Realizing the potential impact of AI, we need to continue to push forward to explore how human jobs, autonomy, and relations are altered in the AI-infused organization (Christin, 2017; Frey & Osborne, 2017; Kellogg et al., 2020; Waardenburg et al., 2022). Such exploration demands a sociotechnical perspective and an increased understanding of how key constructs like trust and distrust (Fulmer & Gelfand, 2012; Lewicki et al., 1998) shape development.

Our study revealed that in the organizational context, continuously introducing a multitude of algorithms, borders between social and technical domains are increasingly fluid and distrust can form an intricate web between social and technical trust referents. We contributed to the IS literature by demonstrating how distrust can spiral when organizations strive to become AI-driven by presenting a model of AI-related distrust dynamics and specifically distrust transference between social and technical trust referents in the organization.

12. References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.

Bachmann, R., Gillespie, N., & Priem, R. (2015). Repairing trust in organizations and institutions: Toward a conceptual framework. Organization Studies, 36(9), 1123–1142.

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2014). Trust transfer in the continued usage of public e-services. Information & Management, 51(6), 627–640.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS Quarterly, 45(3), 1433–1450.

Bijlsma-Frankema, K., Sitkin, S. B., & Weibel, A. (2015). Distrust in the balance: The emergence and development of intergroup distrust in a court of law. Organization Science, 26(4), 1018–1039.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662–679.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. Social Problems, 68(3), 608–624.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies* (First Edition). W. W. Norton & Company.

Chen, X., Huang, Q., Davison, R. M., & Hua, Z. (2015). What Drives Trust Transfer? The Moderating Roles of Seller-Specific and General Institutional Mechanisms. International Journal of Electronic Commerce, 20(2), 261–289.

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. Big Data & Society, 4(2), 205395171771885.

Danks, D., & London, A. J. (2017, August). Algorithmic bias in autonomous systems. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (4691-4697).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126.

Dimoka. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging Study. MIS Quarterly, 34(2), 373.

Dirks, K. T., & de Jong, B. (2022). Trust within the workplace: A review of two waves of research and a glimpse of the third. Annual Review of Organizational Psychology and Organizational Behavior, 9(1), 247–276.

Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. The Academy of Management Review, 23(3), 601.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62–70.

Faraj, S., Renno, W., & Bhardwaj, A. (2022). AI and uncertainty in organizing. In M. A. Griffin & G. Grote (Eds.), *The Oxford Handbook of Uncertainty Management in Work Organizations*, C4.S1-C4.S15. Oxford University Press.

Fountaine, T., McCarthy, B., & Saleh, T. (2019, July). Building the AI-powered organization. Harvard Business Review, July-August 2019, pp.62–73.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280.

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. Journal of Management, 38(4), 1167–1230.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660.

Gong, X., Zhang, K. Z. K., Chen, C., Cheung, C. M. K., & Lee, M. K. O. (2020). What drives trust transfer from web to mobile payment services? The dual effects of perceived entitativity. Information & Management, 57(7), 103250.

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems, 29(2), 101614.

Gustafsson, S., Gillespie, N., Searle, R., Hope Hailey, V., & Dietz, G. (2021). Preserving organizational trust during disruption. Organization Studies, 42(9), 1409–1433.

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.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. Human Factors: The Journal of the Human Factors and Ergonomics Society, 57(3), 407–434.

Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: prerequisites, causes and goals of human trust in AI. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 624–635.

Kähkönen, T., Blomqvist, K., Gillespie, N., & Vanhala, M. (2021). Employee trust repair: A systematic review of 20 years of empirical research and future research directions. Journal of Business Research, 130, 98–109.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 14(1), 366–410.

Komiak, S., & Benbasat, I. (2008). A Two-Process View of Trust and Distrust Building in Recommendation Agents: A Process-Tracing Study. Journal of the Association for Information Systems, 9(12), 727–747.

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. Journal of the Association for Information Systems, 16(10), 880–918.

Leonardi, P. M., Barley, W. C., & Woo, D. (2022). Why should I trust your model? How to successfully enroll digital models for innovation. Innovation, 24(1), 47–64.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review, 23(3), 438–458.

Lockey, S., Gillespie, N., Holm, D., & Asadi Someh, I. (January 2021). A review of trust in artificial intelligence: Challenges, vulnerabilities, and future directions. Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS).

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103.

Lumineau, F., & Schilke, O. (2018). Trust development across levels of analysis: An embedded-agency perspective. Journal of Trust Research, 8(2), 238–248.

Lumineau, F., Schilke, O., & Wang, W. (2022). Organizational trust in the age of the fourth industrial revolution: Shifts in the form, production, and targets of trust. Journal of Management Inquiry, Forthcoming.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.

McEvily, B., Perrone, V., & Zaheer, A. (2003). Trust as an organizing principle. Organization Science, 14(1), 91–103.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. Information Systems Journal, 16(4), 317–342.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. Organization Science, 32(1), 18–41.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Towards human AI hybrids. MIS Quarterly, 43(1), iii–ix.

Renner, M., Lins, S., Söllner, M., Thiebes, S., & Sunyaev, A. (January 2022). Understanding the necessary conditions of multi-source trust transfer in artificial intelligence. Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS).

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3), 393–404.

Sachs, S. E. (2020). The algorithm at work? Explanation and repair in the enactment of similarity in art data. Information, Communication & Society, 23(11), 1689–1705.

Saunders, M. N., Dietz, G., & Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? Human Relations, 67(6), 639–665.

Shao, Z., Zhang, L., Brown, S. A., & Zhao, T. (2022). Understanding users' trust transfer mechanism in a blockchain-enabled platform: A mixed methods study. Decision Support Systems, 155, 113716

Sørensen, O. H., Hasle, P., & Pejtersen, J. H. (2011). Trust relations in management of change. Scandinavian Journal of Management, 27(4), 405–417.

Stewart, K. J. (2003). Trust transfer on the world wide web. Organization Science, 14(1), 5–17.

Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. European Journal of Information Systems, 25(3), 274–287.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. MIS Quarterly, 45(3), 1557–1580.

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. Academy of Management Discoveries, 4(4), 404–409.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. Organization Science, 33(1), 59–82.