Trust in data engineering: reflection, framework, and evaluation methodology

Sarah Oppold[1,*], Melanie Herschel[1]

[1]University of Stuttgart, Germany

Abstract

Trust is and has been essential to human interactions. With the rise of technology, we now live in a socio-technical environment where people frequently interact with technology as well. It is therefore natural to expect that people will also develop trust in technology. Data engineering researchers have at least assumed this when claiming certain methods they devise (e.g, explanations using provenance), likely help to foster some notion of trust. But rarely is the notion of trust clarified or this claim validated. We propose a more systematic consideration of trust in data engineering technology, compared to the ad-hoc state of the art. Therefore, we first review the notion of trust established in other disciplines, based on which we derive a model for trust in data engineering technology. We then present guidelines on how to proceed to devise a trust strategy aiming at enriching data engineering technology such that it potentially fosters trust conforming to our model. We further discuss how to possibly evaluate a trust strategy. We apply our trust model on a use case, for which we devise, implement, and evaluate a trust strategy using our proposed guidelines and methods. The results of our evaluation indicate that statements like "transparency helps build trust" should be used cautiously. This highlights the need for contributions like those we present here, as only a more systematic approach to defining, integrating, and evaluating trust in data engineering can bring us a step closer to provably fostering trust in such technologies.

1. Introduction

Our society depends on us humans trusting each other. From crossing the streets, to collaborating with coworkers, to being treated by doctors, our society is built on trust. The rise of technology and its integration into our world, has created a socio-technical environment where humans live together with technology. This means that we now not only have to trust other humans, we have to also establish a similar relationship to technology rather than second guessing its every "action", in order to benefit, for instance, from its improvements in efficiency or productivity.

In an increasingly data-driven world, data engineering, data analysis, and machine learning are software technologies that can significantly affect human lives (e.g., [1, 2, 3]) and for which some notion of trust has been recognized as an aspect to consider (e.g., [4, 5, 6]). This paper focuses on trust in *data engineering* that encompasses the full data preparation pipeline to get from raw data (as collected) to data "fit for analysis", e.g., data used for training machine learning models. Typical data engineering steps include data transformation [7], cleaning [8], and integration [9]. Data engineering is usually required in any data-driven process and a plethora of systems and algorithms for it exist.

While trust in such engineered data has recently gained attention – yielding approaches to possibly quantify, assess, or even improve trust – we observe that the notion of trust is usually not well defined and does not correspond to the concept of trust established in other disciplines, e.g., philosophy or psychology. In a first line of research, the notion of trust considered in the context of data engineering and data analysis reduces to a possibly related metric and trust in the broader sense is neither considered nor evaluated. For instance, trust as understood in [5] reduces back to the accuracy of a machine learning model. In [10, 11, 12], trust is quantified, e.g., based on the similarity of information and source provenance provided by different data sources. While the resulting trust scores are measured in different settings, it is never validated whether or not the scores actually correspond to some established notion of trust. A second line of research considers transparency and explanations to foster trust (see, e.g., [13, 14, 15]). In this context, data provenance [16], which offers transparency in data engineering pipelines, is frequently named as relevant for evaluating trust (e.g., in [10, 17, 18, 19]). Yet, we are not aware of any validation of this claim. In that sense, the use of the term trust in data engineering has been mostly ad-hoc, without a clear or consistent definition. Furthermore, methods to evaluate solutions for trust in data engineering with respect to such a definition are lacking.

Clearly, we need a more nuanced and systematic discussion on trust in data engineering, to which we contribute considering the following questions: *How can we incorporate the concept of trust into the development process of data engineering pipelines to obtain trustworthy data engineering? How can we assess trust or trustwor-*

Proc. of the First International Workshop on Data Ecosystems (DEco'22), September 5, 2022, Sydney, Australia
*Corresponding author
sarah.oppold@ipvs.uni-stuttgart.de (S. Oppold);
melanie.herschel@ipvs.uni-stuttgart.de (M. Herschel)

thiness in a data engineering pipeline? While we expect there are many different types of solutions, our focus here lies on technical solutions to possibly influence trust in data engineering. Our contributions are: (1) We critically review the term "trust" (Section 2) to define a *theoretical model for trust in data engineering* (Section 3). (2) Based on this model, we describe a framework for *trust engineering* that integrates trust in the data engineering pipeline and serves as a *guideline to develop a trust strategy* (Section 4). (3) We describe a general procedure one can use to *evaluate a trust strategy* (Section 5). (4) We apply our methods to devise a trust strategy to a use case based on a credit scoring scenario, where explanations are integrated into a data engineering step as evidence to possibly foster trust. Our systematic evaluation, however, reveals that the explanations may not reach this goal, highlighting the importance of a more systematic study of the problem with the methods we propose in this paper (Section 6). Note that we are aware that it is possible to manipulate and deceive people by creating an illusion of trustworthy data engineering solutions [20] and that our contributions can lead to such deceptions and manipulations. Countering or regulating this is however out of the scope of this paper.

2. Trust perspectives

As we motivated above, trust in data engineering and analysis has been considered in an ad-hoc manner, while it has been systematically discussed in other disciplines, leading to some common understanding what trust typically entails.

2.1. Philosophical perspective on trust

The discussion on trust has a long history in philosophy [21] and while the concept remains elusive, there are some underlying ideas that most philosophers seem to agree upon. One key facet of the discussion that we highlight here is the distinction of *trust* and its related concept *reliance*. Note that while most philosophical research has dealt with interpersonal trust, our discussion will also review the philosophical perspective on trust in technology.

2.1.1. Reliance

In general, person A *relies* on a proposition p (e.g., that another person performs a certain action) to achieve their goals, when p is a productive means to achieve their goals and p has to be true for its success [22]. Reasons for reliance are often of pragmatic nature [22]. We rely on forces beyond our control or even our comprehension [23].

2.1.2. Trust

Considering trust, a truster A usually trusts a trustee B to do C [24]. As natural, familiar, and elemental it is to trust for us as humans, as complicated it is to describe it as a concept. What philosophers agree on is that trust entails that (1) A is somehow vulnerable to a risk when they trust B, and (2) A relies on B to both be competent and willing to do C [24]. Related to the psychological attitude of trust is the property "trustworthiness" that we can ascribe to others when we think that we can trust them (i.e., we think that they fulfill point 2). While philosophers thus agree that trust is based on reliance (see point 2), they cannot agree what the additional factor is that differentiates trust from mere reliance. While some argue that the trustee's motive must be of some moral nature such as self-interest, goodwill, or moral integrity, others argue that the additional factor is some sort of normative expectation the truster has vis a vis the trustee. It seems to depend highly on the trust relationship example used. A different stance philosophers use to differentiate between trust and reliance is that if B fails A in a reliance relationship, A feels disappointed, whereas in a trust relationship, A feels betrayed [24]. Important characteristics of trust are pro-attitude (truster wants trustee to succeed in doing C), vulnerability, lack of control, and active acceptance of risk [25, 24].

While trust remains an elusive concept, a widely adopted model is the ABI trust model [26]. It identifies three factors of perceived trustworthiness: *(A)bility*, that is the skills or competencies of the trustee, *(B)enevolence*, which refers to the extent to which the trustee is well meaning to the trustor, and *(I)ntegrity*, which is that the trustee seems upright in the eyes of the truster because they share a common set of values or principles. As we shall see in Section 3, we incorporate the ABI characteristics in our trust model targeting data engineering technology rather than a human as trustee.

2.1.3. Trust in technology

While philosophers have studied different variants of trust (e.g., self-trust, trust in groups, trust in organizations), they all are based on human interaction and communication [24]. Technology strongly differs from humans. On the one hand, it lacks human characteristics such as intentionality and hope [27], it cannot use language, and is not free to act as it will [28]. On the other hand, it presents other non-human characteristics such as opaqueness to the user, or unnoticed updates [27]. So can technology be a trustee in a trust relationship according to the previously described notion of trust? Indeed, when people talk about trusting technology, they sometimes talk about a computer artefact, a mere object that is just expected to work as intended, an object that is an

instrument to achieve one's goals. This would be considered what philosophers call "trust as reliance" [27, 28] and not "real" trust.

However, if we take a closer look, technology often is more than just a simple artefact. Technology can feature "logical complexity, capacity to store and manipulate data, potential for sophisticated interaction with humans" [27] and can show unpredictable behavior [27, 28]. Thus, technology seems to encompass more than just mere objects that we rely on. In addition to that, humans, as the partner in a trust relationship with technology, can become emotionally involved in the relationship because trust comes easily for humans [20] who have a capacity to anthropomorphize (form bonds with machines similarly to how they personify pets) [20, 28]. Thus, within a socio-technical system, technology can appear as "quasi-other" with qualities similar enough to humans for them to create a trust relationship [28].

Trust in technology might not be human trust but something similar, lying between interpersonal trust and trust as reliance [27]. It might even be on a spectrum ranging from simple machines that only afford reliance and where the trust is based on functional criteria up to complex autonomous machines with unpredictable behavior that cannot be verified but have to be trusted [20, 27, 28]. Further layers of trust need to be placed in the developers, designers, and company [20], which makes an analysis of trust in technology even more challenging. To make the distinction between trust in technology and interpersonal trust more explicit, researchers have introduced some additional naming and have begun a differentiated discussion. Grodzinsky et al. [27] for example introduce new terms: they call trust in electronic and trust in physical (face-to-face) encounters E-Trust and P-Trust, respectively. Sullins [20] defines different situations of robotic trust, and Coeckelbergh [28] analyzes the impact of different cultures on trust in robots. In this context, our work focuses on E-Trust but rather than focusing on robots, we focus on data engineering technology as trustee.

2.2. Psychological perspective on trust

While the philosophical approach is fueled by the intention to analyze human phenomena, psychologists attempt to assess why we engage in this behavior of trusting or distrusting another person. Psychologists also struggle to conceptualize and operationalize trust behavior, but see the same main characteristics of vulnerability, risk, uncertainty, and pro-attitude that are present in the philosophers' view [29, 30, 31, 32]. We consider psychological studies on behavioral causes to not be directly relevant to the development of a first model of trust in data engineering technology and thus leave their discussion deliberately short.

2.3. Computer science perspective on trust

Finally, we review the perspective from computer science on trust, with a special focus on trust with respect to data processing software that performs or relies on data engineering or data analysis. While trust is also considered in other branches of computer science (e.g., security and privacy), we do not review these in detail due to space constraints.

As we have pointed out in the introduction, the term trust is typically used in an ad-hoc way, yielding different notions of so-called trust that do not necessarily correspond to the common notion philosophy or psychology agree on. In particular, we observe that trust often reduces back to a measurable metric that is indicative of the quality or performance of a solution, but where it is unclear if and how it correlates with trust. Other work advocates that transparency and explanations are key factors to establish trust, which is typically not evaluated or validated though.

2.3.1. Metric-reduced trust

First approaches have emerged to quantify, assess, or even improve what the authors call trust in data processing. For example, [5, 33] attempt to measure trust of machine learning predictions. However, their trust boils down to the precision or accuracy of machine learning results. Similarity-metrics are another category of metrics standing in for trust. For instance, [10, 11, 12] quantify trust based on the similarity of information and source provenance provided by different data sources. While the proposed methods are certainly valuable to improve the likelihood that approaches return the "correct" result and improve the overall quality or performance, this notion of trust does clearly not bear the same characteristics as trust reviewed in the previous subsections.

2.3.2. Transparency and explanations for trust

Several works discuss interpretability and explanations for machine learning models, seen as a possible means to improve trust (e.g., [34, 35]). The general argument is that such methods offer evidence and verifiability that foster trust in a user or developer. Ribeiro et al. [35] evaluate their methods for trust, but this evaluation either simulates users or equates trust with which model performs better (relating back to the metric-reduced trust). Transparency and explanations in data engineering pipelines can be achieved via data provenance [16]. Also in this area, these are frequently named as relevant for evaluating trust (e.g., in [10, 17, 18, 19]). Yet, we are not aware of any work that has studied or validated how transparency and explanations truly relate to trust.

2.3.3. Towards trust modeling

As a starting point to address the aforementioned shortcomings, a more nuanced discussion about trust has recently emerged in the area of computer science. Siau and Wang [15] for example discuss trust in artificial intelligence (AI). They collect a set of different definitions for trust and derive a set of factors for trust in AI technology along multiple dimensions. They also list a variety of approaches to build and then nurture trust in AI. Having focused on methods for trust in AI, this work lacks a catalog of methods for trust for data engineering. Furthermore, it does not include an actionable process taking up their discussion to "implement" trust in AI.

Meeßen et al. [36] derive a model for trust in Management Information Systems (MIS) based on both the ABI trust model [26], which we reviewed in Section 2.1.2, and research in automation and organizations. They translate the ABI terms from interpersonal trust to trust in MIS, allowing a more differentiated discussion about trust in technology. While MIS cover data engineering applications, the proposed trust model is centered around the trustors, mainly identifying factors such as perceived trustworthiness that lead to their use of an MIS. In addition, this work does not model or show what developers of MIS can actually do to build and foster trust that can lead to the decision to use the system.

Thornton et al. [37] call for a more nuanced discussion on the methods developers can use in order to foster trust, proposing what they call *trust affordances*: "characteristics of the technology by virtue of itself or of features designed into the technology to promote trust by providing access to evidence of (dis)trustworthiness specific to a user, a technology, and their context". As they consider technology in a broad sense, the discussion remains very general. We build on their methodology and general ideas to devise guidelines for built-in trust in data engineering.

3. Trust in Data Engineering

We build on the research presented in the previous section to define a trust model for data engineering technology.

3.1. Desiderata

The following desiderata, derived from our discussion of different trust perspectives, underly our model of trust:

- *Distinguishing trust vs. reliance.* The model should incorporate distinctive features that capture trust as opposed to mere reliance. This distinction usually implies the truster's risk awareness with respect to the trustee.
- *Modeling both main parties involved in a trust relationship.* While classical trust models assume both parties to be humans and thus having similar properties, in our setting, the truster and the trustee are inherently different types of entities. Modeling both in detail opens opportunities for a more detailed discussion of what trust in this kind of relationship entails.
- *Modeling influencing factors.* Various factors may influence the kind of trust relationship established between a truster and a trustee, making a concise and unique definition of trust difficult (see Section 2). The model should integrate influencing factors to reflect this ambiguity and incorporate the different nuances of trust, thereby offering a more detailed model for a systematic and multi-facetted study.

3.2. Model for trust in data engineering

Given the desiderata described above, we build our novel model for trust in data engineering. An overview of the model is depicted in Figure 1. Note that it is based on the ABI model [26] discussed in Section 2.1.2, similarly to [36]. While our model is more comprehensive than previous work and tailored to data engineering, we do not claim completeness (it can be extended) and leave open the discussion how far it applies beyond data engineering (our area of expertise).

3.2.1. The truster - a human

In the trust relationship we consider, a *human* is the *truster*. Based on the general notion of trust (see Section 2), we define the human in a trust in data engineering relationship has to be *aware* of a *vulnerability* to some sort of *risk* when using the data engineering technology. Otherwise, the human will use the application as just another tool and we are looking at a "trust as reliance" situation. A human could for example feel vulnerable and at risk when, while using a website, they are aware that they thereby may indirectly divulge preferences or personal information that can affect what information they will be shown, e.g., which news or which job advertisements are recommended. We argue that humans also feel vulnerability when it is not themselves but other people that are subjected to a risk from the trustee.

The trust relationship a truster may or may not engage in inherently depends on several *influencing factors*: The human could be in the *role* of a user of the technology, but also others, such as an examiner, operator, executer, etc. [38]. This will influence how the truster approaches the trust relationship. Humans' decisions to trust are not only influenced by their role, but also by their general *disposition to trust*, their *past experiences* in general (e.g.,

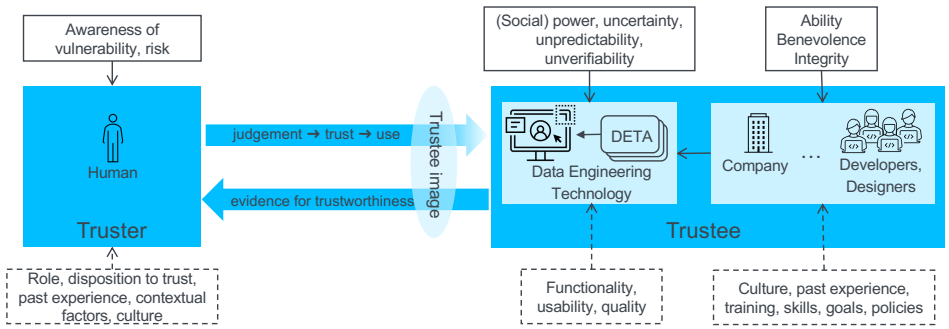

Figure 1: Model for trust in data engineering. A human truster builds a trust relationship with a trustee, i.e., a data engineering application. The latter divides into DETAs and relates to further trust entities (e.g., company). Solid boxes surround necessary characteristics of either the truster or (parts of) the trustee to establish a trust relationship. Dashed boxes group influencing factors.

based on their privileges and power) and in particular with (similar) technology, and *contextual factors* of the interaction. A human's actions are also influenced by the *culture*(s) the human is part of, shaping expectations, behaviors and beliefs [39].

Note that given the large variety of human trusters resulting from different influencing factors and degree of risk awareness, the trust relationship to a trustee can be significantly different from one human to another. For instance, one human's relationship with the trustee may actually be based on reliance because they do not see nor are aware of any risks involved in interacting with the trustee. At the other side of the spectrum, someone else might not engage in a trust relationship at all because they feel too vulnerable and thus decide not to use the system.

3.2.2. The trustee - a data engineering technology

Given the context of our work, the *trustee* is some *data engineering technology*. For the truster to feel vulnerable, it has to have some (social) *power*, element of *uncertainty*, *unpredictability*, or *unverifiability*, thus preventing the assertion that the data engineering technology will not cause any harm.

Typically, such an application is complex and consists of multiple different *data engineering technology artifacts (DETAs)*. These include for instance services, datasets, or algorithms. Note that the truster may or may not be aware of DETAs. Each DETA, as well as the data engineering technology perceived as a whole, is characterized by its *functionality*, *usability*, and *quality*. These have to be sufficient in order for the truster to perceive the technology as reliable. Each DETA could also carry the potential to harm and therefore could also be individually trusted or distrusted by the truster.

Given that technology is shaped by humans and organizations, parties like *developers*, *designers*, or *companies* are part of the trustee in a trust in technology relationship. Note that these are parties with which the truster can also engage individual trust relationships. However, we also include these in the model of trust with respect to data engineering software, because their characteristics can influence this trust relationship as well. Indeed, their ability, benevolence, and integrity have shaped the data engineering technology and can indicate to the truster whether the trustee is trustworthy or not. How parties behind the technology act when developing the product is again influenced by their *culture* - including organizational and functional culture [39] - but also their *past experiences*, *training*, *skills*, *goals*, and *policies*. All of this can affect the trustworthiness of the product, i.e., the data engineering technology, and may be taken into account by the truster when making the decision whether or not to trust the data engineering technology.

3.2.3. Interactions.

We now describe the interaction of the two parties involved in establishing a trust relationship.

When a truster judges the trustworthiness of someone, they are actually assessing pieces of evidence they are provided with to evaluate whether it is worth taking the risk to trust the other party and be vulnerable in some aspect. Whether we are in the process of judging humans or now data engineering technology, we think the human truster continues to act the same. Therefore, we adapt the ABI framework by Mayer et al. [26] (Section 2) which states that the trustee is assessed with respect to their ability (i.e., skills and competences) to fulfil their tasks, their benevolence towards the trustee, and their integrity of principles they act upon. Wile these are classically characteristics of persons and organizations, in our setting, the truster usually creates an imaginary *image* of the trustee based on visuals and communication with the data engineering technology. Indeed, communication to developers or the company behind the application, or access to the codebase are usually not available to the

truster, so their ABI characteristics are transposed to the image of the data engineering technology. Based on the truster's epistemic and practical judgment, the truster then decides whether to trust and then potentially use the technology [36].

Going from the trustee to the truster, the trustee provides evidence towards the truster. In case of data engineering applications, this could be through a modern or old-looking visual interface, whether questions are answered in an FAQ, etc. Opposed to interpersonal trust, trust in data engineering technology involves trust in a complex system of people, groups, institutions, who often cannot be judged directly but only through the pieces of technology the truster has access to. In addition to that, the truster often does not have the capabilities to understand the inner workings of the technology they are supposed to assess. Following the ABI model [26], information on ability, benevolence, and integrity of the trustee with respect to the potential risk might be evidence that increases the perceived trustworthiness.

4. Design data processing for trust

Clearly, when developing data engineering technology, the evidence that can be provided is under the trustee's control, who can adapt this evidence to potentially influence the trust relationship. We propose guidelines on how to systematically integrate trust in the development of data engineering pipelines, by enriching the general data engineering process with further steps fostering trust.

4.1. Assumptions

To align with the trust model we defined in Section 3, we make the following assumptions. First, to guarantee that we are fostering a trust relationship conforming to our model, we assume that the truster is aware technology is used, that it poses a risk to themselves or others, and its functionality cannot be completely verified. Second, we assume that the truster has an ambivalent attitude towards the data engineering technology and can be led to trust it. Finally, we acknowledge that the actions of developers and companies can also create an illusion of trustworthiness, e.g., through clever designed evidence. Here, we assume a benevolent trustee, who intends to provide actual evidence of trustworthiness and does not want to trick the user into trusting a non-trustworthy technology.

4.2. Trust-integrated data engineering

With these underlying assumptions, we enrich the general data engineering process to integrate trust in the technology as summarized in Figure 2. The top of the figure shows the different steps of the data engineering process, whereas the two bottom components "accompany" the whole process from a technical and organizational perspective, respectively.

In general, before developing actual data engineering software, the *goals* to reach with the use of data need to be defined. Based on these goals, relevant data need to be identified and *collected*. As these data may come in various formats from different sources, *data wrangling* is implemented to transform, integrate, and clean the data to obtain a unified and consistent view of the data relevant to the goal. These data can be further *enriched* with application specific data and annotations, before they are *distributed* to downstream data consuming applications such as data analysis techniques. To monitor, document, and support the process, *metadata* are typically gathered and maintained. In addition, a data engineering process is usually subjected to some form of *governance*.

Following our model of trust in data engineering, the data engineering technology in its role of trustee can support a trust relationship by providing appropriate *evidence*. This may involve evidence collected at all stages of data engineering. The methods applicable to collect evidence possibly vary from one stage to another, making it important and challenging to select appropriate methods. The collected evidence can be managed within the metadata management component. While there are many ways to possibly foster trust in data engineering applications, as well as trust in the parties behind the applications that can also have an effect on the considered trust relationship, this paper focuses on the technical solutions targeting trust, leaving the study of trust with respect to governance to the future. This paper also does not aim at exhaustively reviewing how to collect and manage evidence (we mentioned some approaches in Section 2), as for different trust scenarios, different solutions apply or may need adaptation. Instead, our work here offers guidelines on how to generally proceed to systematically integrate the consideration of trust in data engineering technology. This naturally integrates into the conceptual planning phase of data engineering processes (i.e., the leftmost step in Figure 2).

4.3. Identify trust scenarios

Our model for trust in data engineering represents a multitude of scenarios in which humans with specific roles, risks, and vulnerabilities are in a trust relationship with a data engineering technology. Specific evidence will be needed - and at the same time enough - for individual trusters to perceive a particular application as trustworthy. Therefore, it makes sense to identify the specific *trust scenarios* anticipated with respect to the application goal, such that that the collection of evidence can be tailored

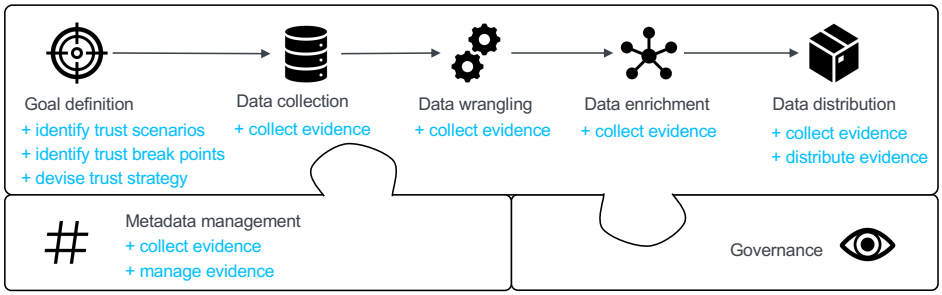

Figure 2: Framework integrating trust in the development of data engineering pipelines. We show the main components of traditional data engineering development in black and our enrichments that integrate trust in blue.

to these.

At this stage, we propose to think about scenarios, relationships, or use cases where the targeted application (goal) has some sort of power over the truster, putting the truster at risk. Modalities of power as identified in the field of political philosophy could be a starting point. Furthermore, different kinds of trusters, i.e., trusters exhibiting different influencing factors, should be considered. It is important to identify which different combinations of influencing factors may define trusters in relevant trust scenarios, as well as the specific risks they potentially face, to then devise trust strategies tailored to the different kinds of trusters. For a wide coverage of possible trust scenarios, we recommend a diverse set of examiners with a critical mindset.

4.4. Identify trust breakpoints

After identifying trust scenarios, it is time to pinpoint the critical parts for perceived trustworthiness in the (planned) data engineering process. We call these *trust breakpoints*. They may comprise methods, algorithms, or other DETAs that could expose a truster to some risk by not meeting specific quality, functionality, or usability guarantees, as their behavior bears some degree of uncertainty, unpredictability, or unverifiability.

It is possible that one trust scenario has multiple trust breakpoints or that different trust scenarios share the same breakpoint. This leads to many-to-many relationships between trust scenarios and trust breakpoints. For each application-relevant combination, we further recommend to determine the requirements each breakpoint in each scenario has to meet in order to minimize or avoid risk.

Since the data engineering software is a technological product, the quality of its trust breakpoints is always shaped by the human capabilities, thoughts, and attitudes of its designers, developers, and surrounding organization. Therefore, there are truster-organization and truster-developer trust relationships to be identified and addressed as well.

4.5. Devise a trust strategy

In a sense, identifying trust scenarios and trust breakpoints can be seen as a requirements analysis on how to cover trust. This analysis forms the foundation to devise a *trust strategy*, i.e., a plan to meet the requirements. Referring back to the distinction of reliance and trust, it will not be enough to provide evidence that convinces the truster that the application is pragmatically the best option to use. Instead, following our trust model, the trust strategy should be designed to provide sufficient evidence on ability, benevolence, and integrity to increase perceived trustworthiness.

The first idea that comes to mind is to transparently provide more information about the trust breakpoints, which the user can use to judge the trustworthiness of the application. This will mostly respond to the ability of the trust breakpoint's DETAs, but could also include evidence for the integrity and benevolence of the company and developers behind the application. Several methods have been developed to provide metadata that can serve as evidence, including plain information about datasets [40], data provenance [16], or machine learning explanations [35]. However, the problem of choosing a suited strategy for requirements given by trust scenarios and breakpoints remains. To systematically devise a strategy and identify pertinent methods, we propose to answer the following six questions in a structured way:

(Q1) What should the trust strategy enable the truster to do? This refers to additional "-ility" requirements of the system that support the truster in their trust assessment and ultimately decision. Answers could include verifiability, reproducibility, traceability [41], reviewability [42], accountability [43], auditability [44], or trialability [45]. Different answers will require different pieces of evidence produced by different methods. For example, verifiability of an output may require an explanation on how the output was generated, whereas the reproducibility of an algorithm asks for information about the algorithm and its parameters.

(Q2) For what kind of component does the truster need

evidence for? Different components of the data engineering technology will require different methods. For example, methods applying to SQL processing [46] significantly differ from methods for data transformations in Map/Reduce pipelines [47]. This question also asks for the granularity of the component that the truster needs evidence of. Whether it is one, multiple, or only the output of a DETA will influence the choice of methods to use.

(Q3) What is the timeframe the truster needs evidence for? Depending on the trust scenario, the evidence should cover past information (e.g., evolution provenance [48]), real-time information (e.g., machine learning model explanation [35]), or future information (e.g., future use of sensitive data [49]).

(Q4) What type of information is needed? To provide the truster with the necessary evidence, different types of information can be used. Examples include factual information such as fairness scores [50], explanations of outcomes [35], or less technical information, e.g., limitations or legal considerations [40].

(Q5) What presentation is appropriate for the truster? Depending on the truster's role, level of expertise, and other characteristics (influencing factors), the evidence has to be prepared and presented accordingly. Therefore, an appropriate level of abstraction and appearance have to be chosen, that provides the evidence without overwhelming the truster. It could for instance be presented like in Datasheets for Datasets [40], where the information is presented as structured text and kept at a very technical level, or the evidence can be presented as in Nutritional Labels for Rankings [50], where the information is (visually) supported using icons, diagrams, and information boxes.

(Q6) What other requirements have to be fulfilled? Since the trust strategy has to fit the overall development plan and requirements, other (technical) requirements may also apply. These could include storage constraints [51], privacy considerations [52], access control [53] or execution speed [54].

After these questions have been answered for all previously determined relevant trust breakpoint-scenario combinations, the developers have enough information to identify or develop appropriate methods.

5. Trust strategy evaluation

After the trust strategy has been defined and implemented, including the collection of evidence, the question remains whether the strategy performs as expected. That is, whether the collected evidence helps trusters to establish a trust relationship with the trustee, in our setting a data engineering technology. In this section, we discuss how the notion of trust we defined in this paper can possibly be evaluated and a trust strategy validated. Given the complexity of human trusters through the number and variety of influencing factors on trust, we postulate that a trust reaction can hardly be simulated, as has been attempted for instance by Ribeiro et al. [35]. Therefore, we suggest to resort to proper user studies, analogously to studies conducted for instance in social sciences or human-computer-interaction, to evaluate a trust strategy. We provide guidelines on how to perform such studies relating to trust in data engineering.

5.1. Study participants and goals

As we have seen, a trust strategy is designed and implemented specifically for a trust scenario. Therefore, the evaluation of the strategy should reuse this scenario in order to validate the strategy with respect to the scenario. This means that participants in the user study should have the same role towards the application as the truster in the scenario. Furthermore, the participants should satisfy the modeled requirements on trusters, i.e., they should be aware that the application is uncertain and its use is related to a specific risk, as defined in the scenario. To ensure this, proper participant selection and gauging questions in the questionnaire of the user study are possible methods one can employ. Additionally, we recommend properly introducing the participants to the scenario, where they should be made aware of their role and the risk the application can pose.

Before deciding on the study setup or devising the questionnaire, the question on what hypotheses to verify needs to be answered. One example of such a hypothesis is: *"The devised trust strategy increases the perceived trustworthiness of the data engineering technology compared to the same technology without trust enrichment."*. Clearly, the hypothesis should explicitly focus on an aspect of the trust model, for which the impact of the trust strategy is then evaluated. The impact itself also encompasses different possible aspects, e.g., perceived trustworthiness (wrt image in the model), actual use, etc. This should be clarified as part of the hypothesis. Finally, the scope of the evaluation needs to be defined, clarifying which aspects of the trustee are covered (e.g., the whole data engineering technology or just selected DETAs).

5.2. Methods for trust evaluation

Once the "what" has been defined, one can address the question on "how" to conduct the study. Here, study designers have to decide which methods to use to evaluate the target aspects. The notion of trust is inherently difficult to quantify, which explains why a set of measurable proxies is usually used that, ideally, highly correlate with the aspects of interest. We review methods that have

been used to evaluate trust and which are amenable to our data engineering setting.

Experiments. For interpersonal trust, researchers have conducted various studies in which the participants could choose between different options [55, 56]. Each of these was implicitly related to trust or distrust based on a risk and reward system. By tracking participants' actions, researches could conclude whether the participants trusted each other or not. This technique can be adjusted for evaluating data engineering technology by creating evaluation scenarios in which the participants can actively choose between different options that correlate with trust or distrust. Recording the decisions of participants can be used as a proxy to measure actual use.

Questionnaires. In designing questionnaires to evaluate trust in data engineering technology, we can adapt and extend questionnaires that have been devised to evaluate trust in other settings. Examples of questions used to measure trust appear in the trust section of the General Social Survey [57] (an annually conducted study in the US). Another option is to derive trust questions analogous to the questions on usability and understandability of the technology acceptance model (TAM) [58]. These techniques allow to examine the thoughts, attitudes, etc. of the participants including perceived trustworthiness, intention to use, and perceived risk.

Structured interviews and unstructured questionnaires. Information about perception, attitudes, etc. that are difficult to express in a question with predefined answers can be collected or captured via interviews or free text fields in questionnaires. This includes, e.g., the reasoning behind participants' answers to structured questions or additional comments on the study. Such answers can provide valuable information on aspects that study designers did not anticipate and offer insights on how to potentially improve the technology, including the trust strategy.

Quantitative metrics. In some settings, it is possible to include quantitative metrics into the trust evaluation. For instance, Wintersberger et al. [59] measure the heart rate of their participants during their study on trust in traffic augmentation for automated driving systems. In their scenario, there was a correlation between heart rate and trust. For data engineering technology, other quantitative metrics such as reaction time may apply.

6. Application of our methods to a use case

After defining our model of trust with respect to data engineering technology as well as guidelines on how to devise and evaluate a corresponding trust strategy, we put our approach to the test by applying it on a real world use case. We describe the use case and its trust strategy development in Section 6.1 and report on its evaluation in Section 6.2.

6.1. Record linkage in a credit scoring application

Credit scores for individuals as provided by companies like Equifax or TransUnion are widely used to evaluate the "creditworthiness" of individuals. This can have a significant impact on human's lives, e.g., depending on their credit score, they may or may not be granted a loan, may have to pay higher or lower interest rates, may be preferred or not in the competitive housing market to sign a lease, etc. Therefore, it is crucial for all parties (the human customers, banks, landlords) that a person's credit history or report, on which the scores are based, is correct and complete. A report itself comprises various customer activities that are shared by different entities (banks, insurances, credit card companies, mobile phone providers, etc.) cooperating with the credit scoring company that are potentially related to the customers' creditworthiness. Examples include opening of a bank account, successfully paying back a loan, etc.

To ensure the data of persons' credit reports are accurate, newly shared customer activities need to be integrated in the consolidated master database of the credit storing company. This is performed by a dedicated data engineering software, which we assume to be similar to the pipeline for a similar goal described in [60]. Following the steps of the general data engineering process outlined in Figure 2, the goal definition is to correctly update the master database, given the data of a newly reported activity record. In this context, the data collection step includes accessing data of the master database (we can assume an SQL query interface) and newly reported records, e.g., obtained via an API. The subsequent data processing that will result in the transformed (updated) master database is all part of data wrangling. Sub-tasks of data wrangling in our use case include the standardization of addresses to all be in the same format, the matching of a record from the master database corresponding to the same person as the new entry (record linkage) possibly followed by human intervention when the match is uncertain (e.g., when no global unique identifier like a social security number is available and not all fields match). If a match is identified, the record on file and the new record are merged to a new record (data fusion). The merged record is then written back to the master database, which can then be queried by subsequent applications, such as an application deriving a credit score.

6.1.1. Trust scenarios

In the use case introduced above, the first step towards devising a trust strategy is to define trust scenarios. To this

end, we first identify various parties (possible trusters) that have some kind of relationship with the data engineering application that can potentially be a trust relationship. These include, for instance, the customers, whose personal data are stored and evaluated by the credit scoring company and the employees of the credit scoring company that should trust the technology to support them in their task of matching and merging records.

Let us now analyze the potential trust relationship between a customer in the role of truster and the data engineering technology (trustee) in more detail. Clearly, the customer relies on the credit reporting technology (e.g., accessible through a web interface) to be able to provide the described service (maintaining the credit report), e.g., to secure a loan. While the customer may be aware of the impact a (wrong) credit history can have on the loan application, the customer usually simply expects the service to work as intended, considering it as an instrument to achieve a goal. As we saw in Section 2.1.3, this rather qualifies as trust as reliance. Also, customers may not be aware that the underlying technology cannot be completely verified and can exhibit quality issues.

This picture changes when we turn our attention to the employees involved in the "human-in-the-loop" data engineering technology as potential trusters in a trust relationship. Clearly, being part of the process, they are well aware that the data engineering technology cannot be completely verified and can cause quality issues. They are also aware of the risk the use of the technology poses, not necessarily to themselves but to their friends, their relatives, and the society in general. For their work, however, they rely on the technology and depending on company policy, the use of the technology bearing some uncertainty with respect to quality may also put these employees at risk, e.g., if, in a performance review it turns out that these employees did match and merge a significant amount of credit reports that have led to claims for correction or to too generous credit scores for non-creditworthy customers. Overall, we see that all criteria are met by employees to be a truster in a trust relationship as defined by our trust model.

On the trustee side, the credit reporting technology comprises several DETAs, e.g., the different steps of the data engineering pipeline we described above. Given the common uses of such technology, it undoubtedly has some social power. As mentioned before, it also exhibits some uncertainty and unverifiability on how the credit reports are generated. Influencing factors relating to the DETAs are mainly their functionality and quality. Besides the credit reporting technology, developers and designers, but also the reporting entities cooperating with the credit scoring company also potentially affect the trust employees put into the trustee.

Given the discussion above, we focus on devising a trust strategy for the trust scenario defined by the trust relationship between the employees and the data engineering technology they use to consolidate credit reports.

6.1.2. Trust breakpoints

For this specific trust scenario we identified above, we consider several trust breakpoints, i.e., DETAs that may affect employees' trust relationship with the technology. A first review reveals for instance that during data collection, trust may be jeopardized by the reporting entities that may transmit erroneous data. During data wrangling, the address standardization may sometimes be inaccurate, depending on which (external) address check service is used. Next, the record linkage may match the wrong records or present the employees with what can be perceived as misleading information to make their decision. Finally, the merge of records could yield an erroneous record. We consider employees unlikely to question the data collection or address standardization DETAs directly (they more likely may not trust external entities serving as data providers, which are other trust relationships). We assume their trust relationship is mostly affected by the internal workings of the assistance the system gives them during record linkage or merge. To demonstrate the development of a trust strategy, we focus on the first of these two breakpoints.

6.1.3. Trust strategy

In order to devise a trust strategy for the trust scenario and breakpoint identified above, we answer the questions proposed in Section 4.5. Essentially, the trust strategy should enable employees of a credit scoring company who consolidate personal data to judge the trustworthiness of technology, which, in this scenario, we assume relates mostly to verifiability of its functionality and quality (Q1). Given the trust breakpoint under consideration, we need evidence for the record linkage component (Q2). As the employees make point-wise match decisions, working with the technology for each individual case, the adequate time frame for evidence is "the now", i.e., real-time (Q3). Considering what type of information is needed as evidence, we argue that developers are probably interested in explanations on how the program came to the conclusion that two records could match, while design decisions on system level and implementations are not pertinent (Q4). In terms of presentation, employees benefit from simple and easy to understand explanations that do not use technical terms from underlying algorithms, as well as visual cues that support the understandability of explanations (Q5). We consider no additional requirements (Q6).

With the answers to the questions given above, we can determine suited methods and algorithms to implement the trust strategy, where we essentially opt to provide

employees with an explanation of matching candidates that serves as evidence of the trustees ABI, so that the employees can potentially gain trust in the system's behavior.

6.2. Evaluating the trust strategy

The goal of the trust strategy in this use case is to foster trust of employees in the data engineering technology they use, by means of explanations. To evaluate if the trust strategy implementation achieves this goal, we conduct a user study, following our discussion in Section 5. This section summarizes the study design, presents results, and discusses these.

6.2.1. Study design

The participants we aim to recruit should take the position of employees of a fictive credit scoring company and review the ambivalent decision of a record linkage DETA. Given the ongoing pandemic, we design an online study. From the different methods for trust evaluation (see Section 5.2), we focus mostly on questionnaires to capture the participants' stance on the data engineering technology. The study includes three main sections, we summarize next. Full details are available on our repeatability website[1].

In its first section, the study provides an introduction to the setting of the study and the topic of record linkage in the context of credit report generation. Thereby, we enable the participants to make informed decisions in the next section focusing on record linkage, and raise their awareness for the underlying potential risk. We further add questions based on 7-grade Likert scales to assess the participants' ambivalent attitude towards the technology they evaluate and their risk awareness with respect to the scenario. Answers to these questions allow us to verify the assumptions stated in Section 4.1. We also include test questions to determine if participants have understood the problem of record linkage.

Next, participants are presented with potential matches, i.e., pairs of records the system suggests to be matches, for which participants, in their role as employees, have to decide if they agree with the system or not. The study comprises 60 matches that each participant reviews. We ensure that these matches cover diverse real-life match situations of varying difficulty in a balanced way. The participants were shown the matches in a random order.

To evaluate the effect of the trust strategy, participants are split into two groups: one gets to see explanations alongside matches, the other group not. Different options for record linkage explanation have been proposed (e.g., [61, 62, 63]). We rely both on the visualization of feature importance by using different color highlights for attributes that are important for making a match decision and attributes that are important towards a non-match. We further provide explanations in the form of human-readable model approximation, listing positive semantic indicators (e.g., important fields firstname, lastname, and date of birth are equal) and negative semantic indicators (e.g., contradictory gender).

In the third section of the study, each participant answers an exit questionnaire that covers several aspects, including usability, by adapting questions from the TAM [58]. We formulate additional questions to assess perceived risk and trustworthiness (see Figure 4), following the same rationale as TAM questions. The answers to these questions again follow a 7-grade Likert scale, ranging from the most positive answer "strongly agree" (1) to the most negative answer "strongly disagree" (7). The study section concludes with a free text field for additional remarks.

During the second section of the study, we capture participants' decision time per match as quantitative metric.

6.2.2. Results

At the time of submission, a total of 19 participants with a computer science background took part in our user study (10 without / 9 with explanations). We opted for participants with a computer science background to ensure all participants have a general understanding of data engineering technology, to better grasp the task we ask them to perform. Based on responses to the first section of the study, we conclude that the participants are generally optimistic that technology can be helpful rather than harmful (mean of 2.7) while they are aware that the technology may put others at risk (mean of 2.7). Thus, they are aware and careful because of associated risks (mean of 2.7).

Determining if the explanations implemented following the devised trust strategy have any effect, we analyze if there is some statistically significant difference between the group of participants without explanations and the group with explanations. Considering reaction time, accuracy of participant match decisions, and the Likert scale questions relating to trust, the applicable statistical tests (t-tests or Wilcoxon-Mann-Whitney-Tests) do not reveal a difference between groups of participants with and without explanations. We thus cannot conclude that explanations have a significant effect on the interaction between employees and the record linkage DETA, in particular, on trust. While the study may benefit from a larger number of participants, the current results show that statements of the sort "explanations are a means to improve trust" should be used cautiously, as it remains an open question in our use case (and others that have not

[1] https://www.ipvs.uni-stuttgart.de/departments/de/research/projects/fat_dss/

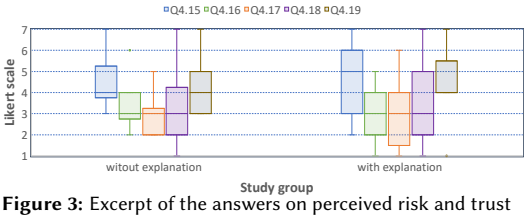

Figure 3: Excerpt of the answers on perceived risk and trust by participants without and with explanations.

Question ID	Question
Q4.15	I would feel safe if people's data were processed by this system.
Q4.16	I would feel at risk if the system was used to decide about me and my data.
Q4.17	I believe in the benefits of the new system.
Q4.18	Assuming I have the power to make decisions in a credit scoring company, I would predict that I would decide to use the system.
Q4.19	I trust the system.

Figure 4: Study questions relating to the judgment, perceived trust, and eventual intention to use

been evaluated), if this holds. Clearly, there is a need for a more systematic consideration of trust, how to possibly integrate it in the design of a data engineering technology (and others), and how to evaluate it. The contributions of this paper are a first step in that direction.

Questions included in the first section and the third section of the study can further be used to compare the "state-of-mind" of participants before and after they have interacted and gained some experience with the record linkage system. Here, we determine that, without explanations, participants show increased trust in potentially risky technology after the study, compared to before the study (p=0.039). This could not be observed in the presence of explanations. On the contrary, we observe a statistically significant decrease in technological optimism and trust for participants that were shown explanations (p=0.009). That is, not only can we not confirm that explanations are helpful to foster trust, but we have an indication that they may actually harm it. A reason may be that explanations give employees further information they can question or that may raise suspicion, outweighing possible benefits of explanations.

While not showing a statistically significant difference between the two studied groups, we still provide some further discussion on the answers to questions relating to the judgment, perceived trust, and eventual intention to use (Q4.14 – Q4.19, summarized in Figure 4). The answers to these questions ranging from 1 (strongly agree) to 7 (strongly disagree) are summarized in Figure 3. Wee see that while the majority of participants in any of the two groups do not feel safe (Q4.15) but rather at risk (Q 4.16), they do believe in the benefits of the system (Q 4.17). Also, the majority of participants, irrespective of whether they have been shown explanations or not, predict they would decide to use the system (Q4.18). However, when directly asking about trust, participants with explanations tend to give a lower rating to perceived trust (Q4.19). Indeed, while all but one participant in this group gave a neutral or negative rating (the median as well as the most positive value are 4), more positive ratings are given by almost half the participants not having seen explanations.

Finally, we report on the two main comments participants provided as part of the final unstructured question. First, participants inquired about further details concerning the step following record linkage, i.e., merging of matched records. This indicates that the second trust breakpoint we identified in our use case is indeed relevant. Second, participants indicated that additional information in the records such as bank account numbers would be helpful for their task. This can be seen as relating to the system's functionality and quality.

7. Conclusion and Outlook

This paper started a nuanced discussion on trust in data engineering. Grounded in established notions of trust from philosophy and psychology, we defined a trust model and proposed guidelines on how to consider such trust when developing data engineering pipelines by devising a trust strategy. Such a strategy ideally fosters trust in data engineering applications, which needs to be validated. To this end, we suggested a general evaluation procedure. We applied our methods to a real-world use case, demonstrating the applicability of the model, guideline, and evaluation procedure. However, our evaluation failed to assert that the explanations we provided as evidence fostered trust in our use case, strengthening us in our initial motivation that statements like "explanations improve trust" may be unfounded. This highlights the need for further investigation on systematically incorporating and evaluating trust in data engineering.

References

[1] J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks, https://propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, 2016.

[2] L. Sweeney, Discrimination in online ad delivery, Queue 11 (2013).

[3] S. Lowry, G. Macpherson, A blot on the profession., British medical journal (Clinical research ed.) 296 (1988) 657—-658.

[4] X. L. Dong, E. Gabrilovich, K. Murphy, V. Dang, W. Horn, C. Lugaresi, S. Sun, W. Zhang, Knowledge-based trust: Estimating the trustworthiness of web sources, Proceedings of the VLDB Endowment 8 (2015) 938–949.

[5] A. Fariha, A. Tiwari, A. Radhakrishna, S. Gulwani, A. Meliou, Conformance constraint discovery: Mea-

suring trust in data-driven systems, in: Proceedings of the 2021 International Conference on Management of Data, 2021, p. 499–512.
[6] X. Zhang, B. Qian, S. Cao, Y. Li, H. Chen, Y. Zheng, I. Davidson, INPREM: an interpretable and trustworthy predictive model for healthcare, in: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2020, pp. 450–460.
[7] P. Vassiliadis, A. Simitsis, S. Skiadopoulos, Conceptual modeling for ETL processes, in: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, 2002, p. 14–21.
[8] I. F. Ilyas, X. Chu, Data cleaning, Morgan & Claypool, 2019.
[9] A. Doan, A. Halevy, Z. Ives, Principles of Data Integration, 2012.
[10] C. Dai, D. Lin, E. Bertino, M. Kantarcioglu, An approach to evaluate data trustworthiness based on data provenance, 2008, pp. 82–98.
[11] C. Dai, H. Lim, E. Bertino, Y. Moon, Assessing the trustworthiness of location data based on provenance, in: 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, 2009, pp. 276–285.
[12] L. D. Santis, M. Scannapieco, T. Catarci, Trusting data quality in cooperative information systems, in: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, 2003, pp. 354–369.
[13] H. Felzmann, E. F. Villaronga, C. Lutz, A. Tamò-Larrieux, Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns, Big Data & Society 6 (2019) 1–14.
[14] M. Janic, J. P. Wijbenga, T. Veugen, Transparency enhancing tools (tets): An overview, in: Third Workshop on Socio-Technical Aspects in Security and Trust, 2013, pp. 18–25.
[15] K. Siau, W. Wang, Building trust in artificial intelligence, machine learning, and robotics, Cutter Business Technology Journal 31 (2018) 47–53.
[16] M. Herschel, R. Diestelkämper, H. Ben Lahmar, A survey on provenance: What for? what form? what from?, The VLDB Journal 26 (2017) 881–906.
[17] B. Glavic, Big data provenance: Challenges and implications for benchmarking, in: Specifying Big Data Benchmarks - First Workshop and Second Workshop, WBDB, Revised Selected Papers, 2012, pp. 72–80.
[18] L. Kot, Tracking personal data use: Provenance and trust, in: Seventh Biennial Conference on Innovative Data Systems Research, 2015, p. 1.
[19] Y. L. Simmhan, B. Plale, D. Gannon, A survey of data provenance in e-science, SIGMOD Rec. 34 (2005) 31–36.
[20] J. P. Sullins, Trust in robots, in: The Routledge Handbook of Trust and Philosophy, Routledge, 2020, pp. 313–325.
[21] Plato, The Republic, 1994. URL: http://classics.mit.edu/Plato/republic.html.
[22] F. M. Alonso, Reasons for reliance, Ethics 126 (2016) 311–338.
[23] M. N. Smith, Reliance, Noûs 44 (2010) 135–157.
[24] C. McLeod, Trust, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Fall 2020 ed., Metaphysics Research Lab, Stanford University, 2020.
[25] J. Simon, The Routledge handbook of trust and philosophy, Routledge, 2020.
[26] R. C. Mayer, J. H. Davis, F. D. Schoorman, An integrative model of organizational trust, Academy of management review 20 (1995) 709–734.
[27] F. Grodzinsky, K. Miller, M. J. Wolf, Trust in artificial agents, in: The Routledge Handbook of Trust and Philosophy, Routledge, 2020, pp. 298–312.
[28] M. Coeckelbergh, Can we trust robots?, Ethics and Information Technology 14 (2011) 53–60.
[29] A. M. Evans, J. I. Krueger, The psychology (and economics) of trust, Social and Personality Psychology Compass 3 (2009) 1003–1017.
[30] J. A. Simpson, Foundations of interpersonal trust, in: Social psychology: Handbook of basic principles, 2007, pp. 587–607.
[31] D. Dunning, D. Fetchenhauer, T. Schlösser, Why people trust: Solved puzzles and open mysteries, Current Directions in Psychological Science 28 (2019) 366–371.
[32] M. Deutsch, Trust and suspicion: Theoretical notes, in: The Resolution of Conflict, 1973, pp. 143–176.
[33] H. Jiang, B. Kim, M. Guan, M. Gupta, To trust or not to trust a classifier, in: Advances in Neural Information Processing Systems, volume 31, 2018, pp. 1–12.
[34] M. Reyes, R. Meier, S. Pereira, C. A. Silva, F.-M. Dahlweid, H. v. Tengg-Kobligk, R. M. Summers, R. Wiest, On the interpretability of artificial intelligence in radiology: Challenges and opportunities, Radiology: Artificial Intelligence 2 (2020) 1–12.
[35] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, p. 1135–1144.
[36] S. M. Meeßen, M. T. Thielsch, G. Hertel, Trust in management information systems (MIS), Zeitschrift für Arbeits- und Organisationspsychologie A&O 64 (2020) 6–16.
[37] L. Thornton, B. Knowles, G. Blair, Fifty shades of grey, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 64–76.
[38] R. Tomsett, D. Braines, D. Harborne, A. Preece,

S. Chakraborty, Interpretable to whom? a role-based model for analyzing interpretable machine learning systems, in: Workshop on Human Interpretability in Machine Learning, 2018, pp. 8–14.

[39] C. B. Gibson, J. A. Manuel, Building trust - effective multicultural communication processes in virtual teams, in: Virtual Teams That Work: Creating Conditions for Virtual Team Effectiveness, Jossey-Bass, 2003, pp. 59–86.

[40] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumeé III, K. Crawford, Datasheets for datasets, in: Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning, 2018, pp. 1–17.

[41] J. A. Kroll, Outlining traceability, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, p. 758–771.

[42] J. Cobbe, M. S. A. Lee, J. Singh, Reviewable automated decision-making: A framework for accountable algorithmic systems, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, p. 598–609.

[43] M. Wieringa, What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, p. 1–18.

[44] R. Cloete, C. Norval, J. Singh, A call for auditable virtual, augmented and mixed reality, in: 26th ACM Symposium on Virtual Reality Software and Technology, 2020, pp. 1–6.

[45] R. Agarwal, J. Prasad, The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies, Decision Sciences 28 (1997) 557–582.

[46] C. Li, Z. Miao, Q. Zeng, B. Glavic, S. Roy, Putting things into context: Rich explanations for query answers using join graphs, in: Proceedings of the 2021 ACM SIGMOD international conference on Management of data, 2021, pp. 1051–1063.

[47] M. Interlandi, K. Shah, S. D. Tetali, M. A. Gulzar, S. Yoo, M. Kim, T. D. Millstein, T. Condie, Titian: Data provenance support in spark, Proceedings of the VLDB Endowment 9 (2015) 216–227.

[48] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, Y. Zhao, Scientific workflow management and the kepler system, Concurrency and Computation: Practice and Experience 18 (2006) 1039–1065.

[49] S. Oppold, M. Herschel, Accountable data analytics start with accountable data: The liquid metadata model., in: ER Forum/Posters/Demos, 2020, pp. 59–72.

[50] K. Yang, J. Stoyanovich, A. Asudeh, B. Howe, H. Jagadish, G. Miklau, A nutritional label for rankings, in: Proceedings of the 2018 International Conference on Management of Data, 2018, p. 1773–1776.

[51] A. P. Chapman, H. V. Jagadish, P. Ramanan, Efficient provenance storage, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, 2008, p. 993–1006.

[52] S. B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen, Y. Chen, On provenance and privacy, in: Proc. of the 14th Intl. Conference on Database Theory, 2011, p. 3–10.

[53] A. Chebotko, S. Lu, S. Chang, F. Fotouhi, P. Yang, Secure abstraction views for scientific workflow provenance querying, IEEE Transactions on Services Computing 3 (2010) 322–337.

[54] N. Bidoit, M. Herschel, A. Tzompanaki, Efficient computation of polynomial explanations of why-not questions, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, p. 713–722.

[55] R. L. Swinth, The establishment of the trust relationship, Journal of conflict resolution 11 (1967) 335–344.

[56] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, C. L. Soutter, Measuring trust, Quarterly Journal of Economics 115 (2000) 811–846.

[57] T. W. Smith, M. Davern, J. Freese, S. L. Morgan, General social surveys, https://gss.norc.org/, 1972-2018.

[58] F. D. Davis, A technology acceptance model for empirically testing new end-user information systems: Theory and results, Ph.D. thesis, Massachusetts Institute of Technology, 1986.

[59] P. Wintersberger, T. von Sawitzky, A.-K. Frison, A. Riener, Traffic augmentation as a means to increase trust in automated driving systems, in: Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter, 2017, pp. 1–7.

[60] M. Weis, F. Naumann, U. Jehle, J. Lufter, H. Schuster, Industry-scale duplicate detection, Proceedings of the VLDB Endowment 1 (2008) 1253–1264.

[61] S. Thirumuruganathan, M. Ouzzani, N. Tang, Explaining entity resolution predictions: Where are we and what needs to be done?, in: Proceedings of the Workshop on Human-In-the-Loop Data Analytics, 2019, pp. 1–6.

[62] A. Ebaid, S. Thirumuruganathan, W. G. Aref, A. K. Elmagarmid, M. Ouzzani, EXPLAINER: entity resolution explanations, in: 35th IEEE International Conference on Data Engineering, 2019, pp. 2000–2003.

[63] S. Gurajada, L. Popa, K. Qian, P. Sen, Learning-based methods with human-in-the-loop for entity resolution, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 2969–2970.