Ethics and Information Technology (2024) 26:37
https://doi.org/10.1007/s10676-024-09774-6

ORIGINAL PAPER

Engineering the trust machine. Aligning the concept of trust in the context of blockchain applications

Eva Pöll[1]

Accepted: 16 May 2024

Abstract
Complex technology has become an essential aspect of everyday life. We rely on technology as part of basic infrastructure and repeatedly for tasks throughout the day. Yet, in many cases the relation surpasses mere reliance and evolves to trust in technology. A new, disruptive technology is blockchain. It claims to introduce trustless relationships among its users, aiming to eliminate the need for trust altogether—even being described as "the trust machine". This paper presents a proposal to adjust the concept of trust in blockchain applications with the tools provided by conceptual engineering. Its goal is to propose a concept of trust, that offers more than a halo term, devoid of the normative depth that the original, philosophical term actually carries. To evaluate possible concepts of trust in blockchain applications, five criteria are proposed: These criteria ensure that the conceptual relation indeed embodies trust, thereby being richer than mere reliance, and that the chosen concept highlights the active role of the trustor. While the concepts of trust in engineers and trust in algorithms have to be discarded, institutional trust emerges as a viable candidate, given some refinements. Ultimately, the concept of trust by normative expectations and attribution is suggested to capture the essence of trust in blockchain applications.

Keywords Conceptual engineering · Trust · Blockchain · Institutional trust · Trust in technology · Blockchain ethics

Introduction

Trust is a basic concept woven into everyday life. Technology is similarly fundamental: it is indispensable, but increasingly complex, which renders it impossible to fully understand all of it. Trust in technology could ameliorate some of the complexity and make it easier to handle. Following this idea that trust can reduce complexity (Luhmann, 1968, p. 25), merely relying on technology expands to trusting it. For example, I trust the word processing program to save my text properly and repeatedly, so I will not lose my progress in case of a program crash—without a deep understanding of the saving intervals, backup implementation, and data structure it uses. And I also trust my bank's online banking software to transfer my money securely to a given recipient. Satoshi Nakamoto, disappointed with the banking infrastructure after the financial crash in 2007, created a new and potentially disruptive paradigm for financial transactions by introducing Bitcoin and hence blockchain technology (Nakamoto, 2008). Blockchain more generally allows the decentralized storage of information in a peer-to-peer network. In the following blockchain technology was said to disrupt not only the financial sector, but was about to change "the nature of social relations and organizations in the global village" (Frizzo-Barker et al., 2020). Some, including Nakamoto, claim that in introducing blockchain the need for trust in particular is overcome; that blockchain allows a transaction system based on "trustless trust" (Hoffman, 2014). This hope, however, is misguided and only withstands a narrow conception of trust, confining it to an interpersonal relationship (Jacobs, 2020). Thus, blockchain applications not only seem to be a disruptive technology, but also conceptually disruptive. A tendency can be observed, that the term "trust" is used without accounting for its further normative implications. Especially the necessity of a vulnerable but active position of the trustor is often neglected, forcing a conceptual misalignment of the concept of trust in the context of blockchain applications. To adjust this misalignment the conceptual structure of trust should be investigated, while keeping the technology of blockchain in mind.

✉ Eva Pöll
eva.poell@uni-muenster.de

[1] Department of Philosophy, University of Münster, Münster, Germany

Published online: 05 June 2024

So far, the research in this area, connecting those on the one hand, who understand the technological aspect of blockchain well and those, on the other hand, who have expertise in ethics and philosophy, is rather thin (Hyrynsalmi et al., 2020; Tang et al., 2019). This research gap results in ambiguous terminology, overestimation, false promises, and unwarranted hype around certain features of the blockchain technology (Jacobs, 2020). This also results in an ambiguous understanding of trust in blockchains. Conceptual engineering could bridge this gap "with the process of assessing and improving our concepts" (Isaac et al., 2022, p. 1), aiming to describe phenomena conceptually correctly and ethically appropriately. To find a suitable concept of trust in blockchain applications I suggest five criteria, along which present concepts of trust in technology are evaluated to find a convincing proposal to adjust the concept of trust in blockchain applications.

Motivating these criteria section two lays the basics of trust and blockchain technology respectively. Based on the predominant paradigm of interpersonal trust (2.a), trust in technology (2.b) is presented. After a brief introduction in the technological basics of blockchain applications, Section 2.c explains how trust in technology generally could be applied to blockchain applications. The main part of this text discusses concept proposals for trust in blockchain applications in Section 3: "Exploring concepts of trust in technology". First, the necessary process to engineer a better concept is introduced, and evaluative criteria are established (3.a). This is followed by the three major concepts of trust in technology found in current research: trust in engineers (3.b), trust in algorithms (3.c), and institutional trust (3.d). As all three conceptions fall short in some way, a fourth conception is offered, building on the concept of institutional trust, but highlighting the active role of the trustor: trust by normative expectations and attribution (3.e). Finally, these considerations are summed up in a conclusion (4).

Trust within the trust-free technology

Interpersonal trust

While this paper does not strive to give a complete outline of the concept of trust itself,[1] a brief introduction seems necessary at this point. Trust is "a form of confidence in another, confidence that the other, despite a capacity to do harm, will do the right thing" (Nissenbaum, 1999, p. 14). It might be the very foundation of society, being present in every interaction, even though it is mostly unrecognized until it is broken (McLeod, 2021). It allows easing cooperation and allows creating goods like meaningful relationships or knowledge, and holds a certain intrinsic value (Baier, 1986; Carter & Simion, 2020). Following this notion, the fundamental concept (and the predominant paradigm) of trust describes the working of an interpersonal relationship (McLeod, 2021). Trust is the attitude of a trustor towards a trustee to act in the way the trustor expects. In short: "A trusts B to φ".[2] The trustor A trusts the trustee B to do or not to do something specific, i.e. to φ. In trusting someone to φ, the following necessary and jointly sufficient conditions are met[3]:

(p_i) *Condition of vulnerability:* rely on the trustee to act according to their expectations to φ;
(p_ii) *Condition of expected competence:* The trustor expects the trustee to be competent to φ;
(p_iii) *Condition of willingness:* The trustor expects the trustee to be also willing to φ; and
(p_iv) *Condition of reliance:* The trustor relies on the trustee to be willing and competent to φ.

Condition (p_i)[4] highlights whether the trustee is free in their decision to act according to the trustor's expectation or not. A trustor's attempt to control the situation, e.g. in minimizing their risk by monitoring the trustee's action, would end the trust relationship (Carter & Simion, 2020). Thus, the trustor remains necessarily vulnerable, usually risking being betrayed or, more abstractly, to lose the value staked at the interaction (McLeod, 2021). Trust is therefore "accepted vulnerability in another's possible but not expected ill will (or lack of good will) towards one" (Baier,

[1] McLeod (2021) provides an extensive account with her entry in the *Stanford Encyclopedia of Philosophy*, so do Carter and Simion (2020) in the *Internet Encyclopedia of Philosophy*.

[2] However, this (trusting someone to do something specific, i.e. a three-place relation) is a thin reading of the concept, compared to trusting someone (in general, i.e. a two-place relation). The latter, thicker understanding seems to demand an interpersonal relationship between agents who exhibit something along (good)will (Baier, 1986) or committent (Hawley, 2014; Lipman, 2023). The thicker reading cannot be widened to encompass also trust in technology, as technology cannot exhibit agency. Thus, for this paper the thinner reading of trust as a three-place relation is of more interest. The considerations of Kelp and Simion (2023) about trustworthiness even seem to suggest that thick trust in someone could be reduced to sufficiently many incidences of thin trust in the same person to do various things. And also others assume the thin reading to be more fundamental (Carter & Simion, 2020).

[3] McLeod states these conditions in course of her text (McLeod, 2021), although she does not reduce them to this essence.

[4] The numbering starts with "p_" to make the later differentiation easier between interpersonal trust conditions (marked with "p_") and technological trust conditions ("t_"). The conditions for trust in technology draw on these conditions for interpersonal trust, thus the similar numbering is chosen to highlight the similarity between the sets of conditions.

1986, p. 235). Conditions (p_ii) and (p_iii) describe competence and performance-related expectations, highlighting that in the view of the trustor it is sufficiently probable that the expected φ-ing is brought about. If the trustor does not believe that the trustee is able to act as they expect (p_ii), or if they do not believe that the trustee will in fact act upon their expectation (p_iii), they would not have (good) reason to trust in the action of the trustee. Finally, condition (p_iv) binds these considerations to an actual action: the trustor not only assumes reasons to trust the trustee with a certain matter but indeed trusts them to act according to the trustor's expectation. Thus, (p_iv) is not inherently implied by (p_ii) and (p_iii). However, it is imperative that (p_ii) and (p_iii) hold true for (p_iv) to become viable. (p_iv) supplements the theoretical conditions established in (p_ii) and (p_iii) with the actual execution of trust by the trustor.

In leaving the action to the trustee, a certain risk remains. Thriving for total security contradicts the concept of trust (O'Neill, 2002). "Trust is important, but it is also dangerous" (McLeod, 2021). Mitigating the risk means placing trust for good reasons, i.e. assessing if the trustee is trustworthy, which "is both epistemically and practically demanding" (O'Neill, 2020, p. 20). Still, as it is part of everyday life, it happens quickly and intuitively and the complexity of this action is easily missed (ibid.). The exact reasons to trust someone to do something are usually unconscious and stating them might even hinder trust (Luhmann, 1968, p. 28). The reasons why the trustee acts accordingly to the trustor's expectations are of lesser interest; it is enough *that* they act accordingly (Luhmann, 1968, p. 23). Others prefer a more concrete understanding, demanding specific reaworthiness, however, is neither necessary nor sufficient for trust itself. It cannot be guaranteed that a trustworthy person is indeed trusted (which is whysons to believe that the trustee will fulfil the expectation, a motive, e.g. encapsulated interest, goodwill, or virtue. This paper follows the thin reading of trustworthiness, that does not commit to a certain motive of action. This also allows a thicker interpretation in the future. Trust condition (p_iv) is relevant) (McLeod, 2021), as trust is granted by the trustor. Thus, trustworthiness is not sufficient for the presence of trust. It is also not necessary to create trust, i.e. one can trust someone to φ, even if the trustee is not trustworthy (e.g. therapeutic trust) or one does not know enough about their trustworthiness (Carter & Simion, 2020). So, trustworthiness is neither necessary nor sufficient for trust, but merely hinting how justified it is to trust someone to φ. Therefore, trustworthiness is not part of the conditions (p_i)–(p_iv) explained above.

Trust in technology

Common language allows stretching the scope of trust beyond an interpersonal relationship, and already includes matters of reliance (Nickel, 2013). Trust seems to be placed in groups like institutions and governments, but also in technology and technological systems (McLeod, 2021; Nickel, 2013). However, not all scholars agree that this is normatively accurate and argue that proper trust can only exist between agents: "People trust people, not technology" (Friedman et al., 2000). The claim is mostly that the term "trust in technology" is too weak and normatively means not more than reliance on an inanimate, but humanized object (Lipman, 2023; Tallant, 2019). So, using "trust" for describing the dependence on technology would be wrong. Others disagree: Ostern (2018) for example, interprets the anthropomorphizing as a signal for the fundamentally social character of the interaction, which allows it to extend the meaning beyond an interpersonal relation. That trusting a technological artefact is a richer attitude than reliance, is shown by the trustor's reaction when an artefact does not exhibit the expected functionality. One does not simply react with curiosity and surprise, if a printer will not print a document or if a word processing program loses the progress of the last two hours. The reaction involves rather frustration and anger—proving a "richer normative attitude" (Nickel, 2013), that is "more than a mere reliability judgment" (ibid.). It is not just an epistemic attitude (which reliance would be), but rather an emotionally and normatively loaded stance, similar to interpersonal trust.

Central for this understanding is the commitment to an apparent and expectable functionality of any given artefact.[5] A certain functionality can be expected based on the design intentions of the manufacturers or engineers, of a suitable use of the artefact and of observing the artefact's performance in the past (Nickel et al., 2010). Thus, "[t]echnical artifacts are in fact embedded in a network of promise" (ibid.). Even though the user (as a trustor) cannot guarantee that the artefact performs according to its functionality, they have control over an adequate context for the technology to perform, e.g. by reading the manual on how to use the software. Even though the functionality of an artefact is usually apparent and its expectation is warranted, there might be cases in which the expectation and the actual performance differ. A main reason might be malfunction, but the artefact could also be poorly designed or assembled, or the intended functionality of the engineers differ from the users' perceived functionality, or those parties differ in their

[5] "Artefact" should be understood in a very broad meaning, including physical objects (like a robot or printer), more or less complex algorithms (like word processing programs and those, which allow secure bank transactions) and technological systems (like a blockchain application and the corresponding infrastructure). In any case it is human-created with an intended, but limited range of functionality and without moral agency or free will (McKnight et al., 2011).

view of values, ethical constraints, and priorities.[6] However, the relevant aspect is the frustration of the trustor's expectation of the functionality, regardless of the origin of the difference between expectation and performance.

Trust in technology builds upon the conception of interpersonal trust, but extends it: the object of trust is no longer a person, with whom the trustor is in direct contact, but a more vague entity. The trustor A places trust in an object B, the technological artefact, that it will φ, i.e. perform according to its apparent functionality. Following McLeod's idea of necessary and jointly sufficient conditions for interpersonal trust (McLeod, 2021), I suggest the following three conditions to describe the demands for trust in technological artefacts[7]:

(t_i) *Condition of vulnerability:* rely on the technological artefact to perform according to their expectations to φ;
(t_ii) *Condition of expected performance:* The trustor expects the technological artefact to be capable and designed to φ; and
(t_iii) *Condition of reliance:* The trustor relies on the technological artefact to φ.

Trust in technology remains to be "a balance between confidence and vulnerability" (Teng, 2021). The trustor is confident that the object of trust will offer the expected functionality (t_ii), while they are still vulnerable in having to rely on another entity (t_i). Yet, the trustor's vulnerability is not due to the trustee's freedom of choice to φ or not to φ, but to the possibility that the technology does not function in the expected way, due to malfunction, misuse or a bad construction on side of the engineers. The trustor remains vulnerable, as they cannot entirely control or guarantee the correct performance of the artefact (especially with more complex technology). Nonetheless, the trustor expects the technology to perform according to its functionality (t_ii). The trustor expects that the artefact is "likely enough to perform in some desired way φ, compared with other alternatives, that it is worth staking something of value on φ" (Nickel, 2013). It should be noted, however, that the technological artefact also is not *competent* to perform according to its functionality, as technology cannot exhibit agency (Lipman, 2023). Technological artefacts cannot intend to do something; they cannot be willing to perform according to the trustor's expectations.[8] Because of its deterministic nature, technology is bound to perform according to its designed functionality, rendering the performance related third condition of interpersonal trust (p_iii) obsolete. Finally, condition (t_iii') anchors the considerations in an actual action. The trustworthiness of technology derives from (epistemic) reasons much like trustworthiness of a person. The reasons encompass positive past experiences of dependability and also involve an assessment of the probability that the artefact will φ, i.e. an evaluation if condition (t_ii) is met. Today, technological artefacts are increasingly complex (Nickel, 2013), thus it is hard for users to sufficiently understand their functionality to evaluate it (Jacobs, 2020; Nickel, 2013; Pesch, 2019). Users take a "leap of faith" (Teng, 2021; also Möllering, 2006) and apply trust to reduce complexity (Luhmann, 1968, p. 28; Teng, 2021).

This does not imply that every interaction with technology must involve some form of trust. Most cases of using a technological artefact, are probably mere reliance. However, if the vulnerability of the trustor (condition (t_i)) is added to the equation, it creates a risk that falls back onto the trustor when the technology does not provide the expected functionality. Similarly, if the trustor is not (yet) willing to actually rely on the technological artefact in practice, i.e. if condition (t_iii') is not met, the user's attitude towards the artefact is also not trust. Only the combination of all three conditions is sufficient to establish trust.

Blockchain in a nutshell

Before the concept of trust in technology is applied to blockchain applications in the next section, a brief introduction to the relevant basics of this disruptive technology should be given. A blockchain is a structure for storing data decentrally (Bundesnetzagentur, 2021). The name "blockchain" originates from the way data is stored: Multiple pending transactions (i.e. data to be stored) are aggregated into blocks that cryptographically refer to their predecessor, whereby they form a chain with the most recent block at the front (Nakamoto, 2008). While conventional databases store data in a central location, a blockchain is redundantly stored in a decentralized computer network. This network consists of three different kinds of actors: participants, passive and active nodes. The passive or full nodes are fundamental: those are servers storing a complete copy of the blockchain.

[6] I want to thank an anonymous reviewer for highlighting that the assumption of malfunction alone is too narrow.

[7] Differences in the phrasing to the original, interpersonal conditions are set in italics. The numbering is prefixed with "t_" to differentiate these conditions for trust in technology from the conditions for interpersonal trust introduced earlier, while still highlighting how close both sets of conditions are.

[8] It is certainly an interesting and ongoing discussion, if some artefacts, like strong AI or robots controlled by a strong AI can be assumed to possess some form of agency. But these considerations go beyond the scope of this paper. Leaving aside this exception, I am confident that technological artefacts cannot be willing to (not) do anything.

Active nodes, the so-called miners, are servers as well and create new blocks and thereby publish information on the blockchain. Participants are users of the blockchain (and thus usually the largest group of actors). They are the actors, who initiate transactions (Bundesnetzagentur, 2021; Schlatt et al., 2016).

Blockchains can be characterized along two dimensions: how limited the access is to the data they contain and whether participation in the administrative process is restricted. In public blockchains, anyone can use the system, i.e. issue new transactions and access existing data. Private blockchains are not freely accessible, but only to be used within an organization or consortium. Typically, private blockchains are permissioned, meaning that a consortium or organization determines who is allowed to run a node in the network and therefore who can create new blocks or store a copy of the blockchain. In order to make the structure of the blockchain as transparent as possible, permissionless blockchains often run on open-source software. That means, that the code, on which the blockchain is running, is developed by individuals who are often only connected through the platform and organized in a decentralized structure. The code is then made public, so that other developers can check and improve it, warranting the correctness of the functionality on the one hand, and the good nature on the other, i.e. that no ulterior motives, biases, or backdoors are brought into the code. Thus, the development is seen as normatively self-regulating within the group of developers, holding each other accountable to normative standards. Even though the individuals might not be trustworthy per se, the blockchain mechanisms connect them in a way that they serve the network as a whole by pursuing their individual interests (Lipman, 2023). While public, permissionless blockchains achieve a maximum in decentralization, private, permissioned blockchains are more centrally controlled, as they are organized around the controlling organization or consortium (Schlatt et al., 2016; Yaga et al., 2018). Thus, a central authority is responsible for the maintenance of the network and the used software, which introduced are more specific and in this possibly more trustworthy point of regulation compared to public permissionless blockchains (Yaga et al., 2018). The transparency of a blockchain therefore also depends on the type of implementation and is greatest for public, permissionless blockchains.

The redundant, decentralized storage of the data also ensures its availability and makes the system resilient to failures of individual nodes (Bundesnetzagentur, 2021). It also leads to a high level of data integrity: As soon as data is stored in the blockchain, it can no longer be changed since it is also stored in all subsequent blocks thanks to the cryptographic value of that block that is referenced in all succeeding blocks. Thus, neither attackers nor authorized persons can modify data once stored, because changing the data would affect all following blocks, demanding to re-instantiate the whole chain from the changed block on. This is unrealistically complex and thus practically impossible (Nakamoto, 2008). These attributes of blockchain systems: transparency, availability or resilience (due to distributed and redundant storage) and data integrity are factors that invite trust in the system (Bundesnetzagentur, 2021; Marella et al., 2020; Tang et al., 2019; Teng, 2023; Völter et al., 2021).

The most relevant activity in using a blockchain application, when it comes to trust, is to issue a transaction.[9] To illustrate this as a matter of trust, the exemplary transfer of an amount of a cryptocurrency (such as Bitcoin) is used: The trustor A would be a user of the blockchain application, i.e. the person who issues a new transaction. Typically, they would not also be miners, but averagely informed and competent consumers with timewise and cognitive limited capacities. Thus, they typically will neither have the expertise, nor will they be able to fully understand the technology, nor invest the time to keep constantly up-to-date with new developments (Teng, 2021; Völter et al., 2021). The user (A) trusts the blockchain application (B) to properly process their data (e.g. the information of a Bitcoin transfer) (φ). To φ should be interpreted as the proper processing of data, i.e. to store the data completely and securely on the blockchain, without diverting, changing or adding information to this transaction, within a reasonable amount of time and for the previously agreed fees. Exactly what the object of trust (B) is, is the matter of the next section. The choice of the characterization of the object of trust will be crucial in deciding on a concept of trust in blockchain applications.

Exploring concepts of trust in technology

Criteria for a concept of trust in blockchain applications

The introduction of blockchain applications put pressure on the concept of trust, as some even claimed it could overcome trust altogether (Hoffman, 2014; Nakamoto, 2008), or artificially create it (as seen in the portrayal of blockchain as 'The Trust Machine', 2015). This caused a conceptual misalignment (cf. Hopster & Löhr, 2023), where the notion of "trust" as currently used in at least part of the research literature may no longer serve its function. The function

[9] Certainly there are more possibilities to form trust, for example in expecting that the miners would be fairly rewarded for investing capacities in creating new blocks; Or the very pressing problem of how to guarantee data quality, i.e. trustworthy data (keyword: 'garbage in, garbage out'-problems). But to limit the scope of this paper and allowing to be specific, I want to focus on the user perspective.

of trust in technology appears to be twofold: firstly, trust allows individuals to handle risky situations. In scenarios where there is limited control, trust enables individuals to take action despite the immanent risk. For instance, when using the banking software provided by my financial institution, I have limited control over ensuring that the money I transfer fully reaches its intended destination. But, given the lack of viable alternatives and the minimal evidence of risk associated with using the software as intended, trusting the banking software instils confidence in my actions. Secondly, trust simplifies interaction. In our modern society, where technology permeates everyday life, it is challenging to avoid using any technology. Yet, as technology becomes increasingly complex, users may lack a deep understanding of its intricacies. As a resort they turn to trusting the technology to function as expected, allowing them to engage with technology without fully comprehending its underlying mechanisms. Trust then simplifies complexity. For example, I trust a word processing program to save my work regularly, even though I cannot explain the algorithm behind this process. Instead, I trust that the program will perform as anticipated, minimizing potential data loss in the event of a program crash.

Blockchain applications were praised as the technology to warrant the risk of trust. Users expected total security, algorithmically ensured, but overlooked that achieving this level of security requires a deep understanding of the technology. But only a small fraction of users possesses this level of understanding, while the vast majority has to fall back on trust in the technology. This fallback might seem innocent, given that blockchain is hailed to be designed to mitigate the risks associated with trust, but in fact it presents conceptual challenges on two fronts. Firstly, by disregarding the relevance of the vulnerability conditions for trust (cf. (p_i), (t_i)), the assumed concept of trust becomes indistinguishable from mere reliance. If the vulnerability condition is indeed irrelevant in the application of the concept, "reliance" should be used instead. However, vulnerability is pertinent in the context of blockchain applications. Users often bear significant responsibility when using blockchain, especially considering the current lack of user-friendly software (Hünseler & Pöll, 2023). Furthermore, with blockchain applications promising availability and resilience, they appear well-suited for storing critical data such as financial records or land register entries (Schlatt et al., 2016). Consequently, there is a high stake associated with the data handled, resulting in a high vulnerability for those who entrust blockchain applications with handling their data. Illustratively, there have been cases of users forgetting their passwords to access their bitcoin accounts and subsequently losing substantial amounts of money (Popper, 2021), and users must be aware of the risks involved in trusting blockchain applications. This is highlighted by the first criterion for the success of this conceptual engineering project: a proper concept of trust should uphold the normative significance of trust and distinguish it from merely relying on technology (summarised below as criterion (a)). The other problematic aspect of the current use of trust is interconnected. Due to the perceived security resulting from the lofty promises of blockchain applications, users tend to overlook the necessity for risk management and seek simplification in the wrong place, hoping for trust through security. A successful concept, however, should accomplish its functions (Thomasson, 2021). Therefore, a successful concept of trust in blockchain applications should reflect the capability to simplify contexts and enable to act under risk and limited control of a situation. In this, the origin of vulnerability and the target of trust needs to be highlighted (cf. criterion (b)).

On the technical side in particular, the term "trust" seems to be portrayed as a dazzling promise of the quality of the technology, but upon closer examination, it remains vague—similar to the use of "trustworthiness" in the context of artificial intelligence.[10] To avert that trust in blockchain technology becomes a mere halo term, the concept needs to be adapted to prevent it from "conceptual degradation" (Hopster & Löhr, 2023). Conceptual engineering should have a descriptive and a prescriptive aspect, capturing how the term is currently used and guiding how it ought to be used (Isaac et al., 2022). Mirroring the descriptive aspect, another criterion for success of the concept adaption process describes that a concept of trust in blockchain applications should be applicable to that particular context (c). The prescriptive aspect is expressed in the conditions to establish trust either as an interpersonal relationship (cf. conditions (p_i)–(p_iv)) or as trust in technology, expressed in conditions (t_i)–(t_iii'). By explicitly grounding the concept in the ethical literature on trust, its normative significance can be enforced. Thus, a fourth criterion of successful conceptual engineering is to align with the necessary and jointly sufficient conditions for trust as an interpersonal relationship or for trust in technology (d).

Another aspect of the trust relation is regularly overlooked, particularly by researchers with a technical background: they often fall for an excessive enthusiasm for technology, assuming that blockchain technology might induce trust on the side of the trustor (see e.g. Becker & Bodó, 2021; Bundesnetzagentur, 2021; Buterin, 2015; Völter et al., 2021). In this they mistakenly reverse the direction of the trust relation by assuming that blockchain technology might induce trust on the side of the trustor. This usage is not only semantically vague but also epistemically deficient, leading

[10] According to Reinhardt (2022), the term "trustworthiness" subsumes everything deemed good in the current research literature about AI.

to incorrect deductions (Isaac et al., 2022). As this fallacy appears to be quite common, a concept of trust in blockchain applications should rectify the epistemic deficit of the mistaken direction of trust, accentuating the central role of the trustor, i.e. the user of the application, and classifying that trust cannot start from the technology side. This is criterion (e) of a successful conceptual engineering project.

In introducing these criteria for a successful conceptual engineering project, also important steps were taken to shape the project (cf. Löhr, 2023a). The disruptor and the disrupted concept were identified: Blockchain applications disrupted the common usage of the concept trust in technology, causing a conceptual misalignment. Many scholars grapple with accurately describing the trust relationship between users and blockchain, with some even suggesting that it has been annulled. This poses a threat to the normative significance of the concept and undermines its functional utility. Furthermore, stakeholders were identified: computer scientists, introducing and promoting the new technology, and ethicists who urge caution and pay attention to the concepts implied. Alongside these stakeholders, users of blockchain applications should be considered as key actors. The disruption blockchain poses on trust seems rather shallow so far: Scientists concerned with the ethical implications of trust in blockchain applications are mostly concerned with the shift in meaning on a rather individual than social level. However, as blockchain gains wider adoption among users, combined with computer scientists' inexperienced use of ethical terms in other research areas,[11] the depth of the disruption could increase.

One goal of this paper is to convince members of the discussion of a more constrained use (cf. Isaac et al., 2022) of the concept "trust" by changing the expectations of actors, targetting the entitlement to apply the concept (cf. Koch et al., 2023). For instance, "trust" should be reserved for describing a relationship involving some kind of vulnerability between the user and the technology, while the concept of reliance should be used in cases where vulnerability is insignificant. To implement this realignment of meaning, the relevant expectations need to be altered through public argumentation (Thomasson, 2021) and the implied norms should be reassessed (Hopster & Löhr, 2023). The corresponding exit rules (cf. Löhr, 2023b) align with the presented criteria for a success of this conceptual engineering project: The use of "trust" needs to be sharpened by explicitly broaden the extension to include blockchain applications, without losing the normative expressiveness of the concept. However, this endeavour must be mindful of the ripple effect, affecting nearby concepts (Löhr, 2023a). The closest concept affected is reliance, against which the concept of trust should be clearly distinguished. By sharpening the border of trust, the concept of reliance may also become clearer. Furthermore, concepts such as transparency, availability, and resilience, which are core values of blockchain applications, may also be influenced upon further investigation. However, given the relatively shallow depth of the disruption, the ripple effect should have limited expansion. Picking up on the proposal of institutional trust in following, also the concept of institution should be examined. It should be answered whether and under which conditions blockchain applications could be considered as institutions.

To find a possible solution how trust in blockchain applications could be modelled, three major concepts are explored to shed light on how they characterize the object of trust and the implications involved: trust in technology, as trust in the engineers behind the artefact, trust in the technological artefact itself, i.e. trust in algorithms, and treating trust in technology analogously to institutional trust. In the following subsections, each concept proposal is evaluated against the criteria of a promising concept of trust in blockchain applications, particularly against criterion (d), which concerns meeting the conditions for trust. To summarise the criteria introduced above, they are listed in short below:

A concept of trust in blockchain applications should

(a) Encompass and explain more than merely relying on technology;
(b) Reflect the functions of trust in technology: simplification and risk management;
(c) Be particularly fitting to the context of blockchain applications;
(d) Align with the necessary and jointly sufficient conditions for interpersonal trust (cf. Section "Interpersonal trust", conditions (p_i)–(p_iv)) or for trust in technology (cf. Section "Trust in technology", conditions (t_i)–(t_iii')); and
(e) Emphasize that the trust relation emanates from the trustor.

Trust in engineers

A first attempt to capture how to trust technology and blockchain applications in particular, stays close to the notion of interpersonal trust and resolves the trustee to be the engineers of a platform. Following a narrow understanding, trust is only possible in the direct interaction of rational and emotional agents. As technology lacks agency and cannot in fact act, trust in technology is actually to be understood as trust in the engineers of the artefact (Fries, 2022; Jacobs, 2020). Trust shifts from a direct, personal interaction to an indirect

[11] As referenced above trustworthiness in context of AI seems to be used rather simplistic, neglecting the normative significance. Also similar ethical concepts like responsibility are under pressure (Himmelreich & Köhler, 2022).

connection qua the technology. The direct connection to a specific trustee is loosened; it is not clear whom exactly one trusts. Instead, trust is placed in a generic group of people, the engineers of the technology (Nickel, 2021). They are expected to act not only technically but also normatively responsible for the developed technology. Thus, they are taking not only care of the technical functionality, but also the moral implications for the user directly (e.g. data security) and for society as a whole (e.g. mitigating subconscious biases). In this, engineers can be trusted to have taken all necessary and sufficient measures to make the technology reliable and safe to use (Nickel, 2021; Teng, 2021).

This understanding of trust, as shifted to the people behind the technology, is fitting to the decentralized structure of blockchain technology. "Engineers" should be understood as a term for those responsible for the correct functionality of a technology, rather than the actual job title. The collective in charge of maintaining the blockchain would be the software engineers, but also miners and node providers. In echo of the blockchain as a means of distributed data storage, some suggest the term "distributed trust" (instead of "trust in engineers" in the context of blockchain applications), to highlight the shift from a direct trust relation to a non-specific collective of blockchain providers and programmers (Jacobs, 2020; Mallard et al., 2014). The conception of distributed trust as an indirect, more generic, but still basically interpersonal relation applies to blockchain applications on two levels: First, it applies to the level of software development (i.e. of software engineers). The code is typically open-source based, enabling the developers to check and improve it, warranting the correctness of the functionality on the one hand, and their good intend on the other hand. Thus, users have to trust this system to be normatively self-regulating within the group of developers. Second, storing data on the blockchain is inherently distributed and involves multiple agents. To establish an information within a block in the chain, it has to be verified and affirmed from numerous nodes of the blockchain network. Users have to be trust that these mining nodes are operated by trustworthy individuals, who do not seek to attack the database and keep the software of the blockchain application up to date. Equally, users also have to trust the providers of passive nodes, to fulfil the users' expectations of maintaining the blockchain (Finck, 2018; Mallard et al., 2014; Teng, 2023; Wang et al., 2022). Trust in blockchain applications "ultimately requires trust in the collectivity of individuals who architect this technology, as well as in the procedures that govern their behaviour and manage their accountability" (Finck, 2018).

Trust in using a blockchain application as distributed trust is actually not trust in technology itself, but still trust in people. Therefore, it has to align with the original conditions suggested by McLeod (2021) to establish interpersonal trust (see Section "Interpersonal trust". (p_i)–(p_iv)) not the conditions particularly for trust in technology. Due to the complexity of the technology, the user is vulnerable and relies on the collective to maintain and provide the blockchain's correct functionality (φ, see Section "Blockchain in a nutshell"). The vulnerability condition (p_i) is met. The conditions of expected competence (p_ii) and willingness to φ (p_iii) demand the trustee to be competent and willing to act on the users' expectations. However, in spite of the self-regulating ideal, in reality, relevant development and maintenance decisions are often not made with the possibility of public influence. Even if these decisions are public the decisions and algorithms are usually only insightful for experts (Teng, 2021, 2023; Walch, 2019). This renders the users more vulnerable and questions if they can estimate competency and willingness. Moreover, developers and providers might have ulterior motives and non-transparent interests (Fries, 2022). Thus, it cannot be secured that the trustee is willing to φ—the condition of willingness (p_iii) cannot be met. The same is true for the condition of expected competence (p_ii): Because the connection between the user and the people behind the technology is highly obfuscated and because (especially in permissionless blockchains) anyone could participate in providing and maintaining the platform, it seems impossible to evaluate the trustee's competence to φ; also the condition of expected competence (p_ii) is failed. To find a proper concept proposal for trust in blockchain applications the conditions to establish trust need to be met. This is expressed in criterion (d) for a successful conceptual engineering project, demanding the proposed concept to align with the conditions for trust (cf. Section "Interpersonal trust"). Consequentially, failing to meet the conditions for interpersonal trust and therefore criterion (d), distributed trust is not a convincing proposal to shape the concept of trust in blockchain applications.

Trust in algorithms

An alternative proposal places the technological artefact in the centre. Instead of trusting engineers indirectly qua the artefact, a direct trust relation between the user and the artefact is modelled: one would trust the functionality of the artefact itself—not merely that someone provided this functionality (Nickel, 2013). This is convincing particularly with regard to the increasing obfuscation of the connection between user, artefact, and the people behind the technology (Nissenbaum, 1999). McKnight et al. argue: "The competence of a person and the functionality of a technology are similar because they represent users' expectations about the trustee's capability" (McKnight et al., 2011). Following this perception, some suggest that proper coding can generate trust as it manifests and reinforces contractual agreements in the form of algorithms. In code expectations, rights, and responsibilities are made explicit, guiding and limiting user

interactions, but also providing some clarity, simplifying the use of ever more complex technology. Normative values (in the case of blockchain, especially transparency, cryptography, and consensus) are encoded and anchored in the algorithm (Bundesnetzagentur, 2021; Finck, 2018; Teng, 2023; Werbach, 2018). As average users are typically unable to fully understand the technical artefact they use and trust, they benefit from trusting it in a reduction of complexity and acknowledging one's epistemic limitations and hence vulnerability (Luhmann, 1968, pp. 14, 96). Understanding trust in technology as trust in the artefact, i.e. algorithms or code, gives up on a motive-based account of trust, and uses the thinner reading, introduced above (see Section "Interpersonal trust") (Jacobs, 2020).

In the beginning of the blockchain movement, it was often postulated that blockchain applications eliminate the need for commonly trusted third parties—even for trust in the first place (Teng, 2021). In the pioneering document "Bitcoin: A Peer-to-Peer Electronic Cash System", blockchain founder Satoshi Nakamoto postulates the replacement of interpersonal trust with algorithms, demanding "cryptographic proof instead of trust" (Nakamoto, 2008). Blockchain was said to introduce "trustless trust" (Hoffman, 2014), being a "trust-free" technology (Lipman, 2023). The blockchain technology is intended to enable users to engage in direct but secure exchange without having to trust a third party (originally the banks that were hit after the financial crisis) or even having to trust directly the other party of a transaction (Buterin, 2015; Hoffman, 2014; Nakamoto, 2008). Trust, which is inherently risky, was meant to be overcome (Buterin, 2015) and, according to the promise, to be automated: Cryptographic signatures ensure the identity of the other party and the origin of the data, while the public visibility of all transactions is to guarantee their correctness (Nakamoto, 2008; Tang et al., 2019). Claiming the defeat of the risk of trusting, presupposes, however, a deep understanding of the functionality and implications of a blockchain application (Jacobs, 2020; Teng, 2021, 2023). Most users are lacking this competence, having neither the time nor the expertise to understand and to keep up with the rapid developments (Lustig & Nardi, 2015). This lack of understanding is factually rendering users unable to ensure their own safety while using an application and thus needing to trust it (Teng, 2021). Blockchain applications, thus, do not enable users to overcome trust.

In fact, "from a user-centred perspective trust is still needed in blockchain-based interactions" (Teng, 2023). Thus, instead of replacing the need for trust completely, most researchers soften the call and claim that trust *is* necessary. Indeed, some third parties can be cut down by blockchain applications (e.g. Bitcoin disintermediates money transfers and eliminates the need for going through classical financial institutions), but the necessity of trust is not eliminated, rather shifted to the working of technology itself. Blockchain applications can be seen as a "trust machine" ('The Trust Machine', 2015), creating trust through code. Most literal is therefore the conception of trust in the blockchain applications as trust in algorithms, i.e. the code of a blockchain application. The code mediates all interactions of users in the blockchain. For example, the algorithms to validate transactions allow to cryptographically prove the authenticity of data, and that it was not altered during the process of storing it. Thus, in adding control of the interaction and security mechanisms the code "may enhance trust" (Smits & Hulstijn, 2020; similarly Marella et al., 2020). Instead of "people trust people" (Friedman et al., 2000) the call is now for "in proof we trust" (Werbach, 2018, p. 29). As "nothing is assumed to be trustworthy […] except the output of the network itself" (Werbach, 2018, p. 29), it is not necessary to form a direct relation between participants (among users and providers distinctly and jointly). Rather, it is enough that each trusts the platform and algorithms to secure the interaction. The contractual conditions of the interaction are cast in code (Finck, 2018, p. 12).

Trust in algorithms tends towards the wish for total security by mitigating all risks through code. Trust is expected to be created through security. Secure algorithms are said to culminate in "trust-creating attributes" (Marella et al., 2020; see also Wingreen & Baglione, 2005). Philosophical accounts of trust in technology, however, criticize this effort as misled and conceptually contradictory: trust necessarily requires vulnerability, not minimized risks, on the side of the trustor (condition of vulnerability, (t_i)). Thriving for eliminating all risks through algorithmic guarantees and proof, in turn, actually prohibits, not creates trust (O'Neill, 2002; Reinhardt, 2022). Furthermore, security, especially in the context of complex technology, is an illusion: securing one area still leaves others open (Nissenbaum, 1999). Trust in algorithms follows the wrong track about vulnerability, thus the condition of vulnerability (t_i) cannot be met. The condition of expected performance (t_ii), however, is met: As a blockchain application is designed to fulfil its functionality, i.e. the users expectations, it can be expected to φ. Even if a blockchain encodes values that invite trust, it is not guaranteed that the trustor would indeed rely on it: the condition of reliance (t_iii') cannot be secured or forced. In not meeting all conditions to establish trust in technology, and thus failing to meet the corresponding criterion (d) of a promising concept, the concept proposal of trust in algorithms falls short in capturing the essence of what it should mean to trust a blockchain to φ. Thus, this proposal should be rejected.

Institutional trust

A third proposal of trust in technology makes use of the conception of trust in institutions. In this the trust relation

is more abstract, as it is neither directed to a specific thing nor a certain person, but to an institution as a whole, including the services it offers through the people representing it, guided by the incorporated rules and incentives (Lahno, 2001; Teng, 2021; Townley & Garfield, 2013). In engaging with an institution (e.g. using a service it offers) the trustor forms predictive and normative expectations, based on three aspects: past experiences with representatives of the institution (Lahno, 2001; Möllering, 2006); the perceived structural assurances (like rules, guidelines, and incentives) promoting specific behaviours (Möllering, 2006; Wingreen & Baglione, 2005) and, connected to those assurances, the perceived values and objectives pursued by the institution (Teng, 2021; Townley & Garfield, 2013). Each interaction between a customer and an institution contributes to shaping the institution's reputation, which offers evidence of its trustworthiness (Alfano & Huijts, 2022; Smits & Hulstijn, 2020). These interactions are mediated by representatives who operate within internal rules, roles, and processes designed to ensure specific outcomes (Lahno, 2001; Wingreen & Baglione, 2005). Or in other words "formal controls, embedded in institutions" (Smits & Hulstijn, 2020). These assurances align with the institution's desired objectives. Institutions take on social responsibility in creating and cultivating social goods (such as knowledge, culture, art, etc.) (Townley & Garfield, 2013) and thereby also incorporate moral values: "institutions can be understood as entities carrying predefined normative qualities, such as moral, social, and legal norms" (Teng, 2021). For example, a university is expected to use its capacities to educate and grow knowledge through research, i.e. create social value, and not sell inventions to malicious military instead, which would imply abandoning the creation of social value and focusing on financial gain (Townley & Garfield, 2013). Social media networks, as another example, are expected to be a neutral platform, on which users can interact securely, without being spied on and sharing their data only with their intended audience. Facebook, however, has repeatedly disappointed these expectations so that "Internet users widely distrust Facebook" (Kelly & Guskin, 2021). Accordingly, the expectations are not only predictive but also evaluative, serving as criteria for assessing, if trust in the institution is warranted (Teng, 2023). The key aspect of institutional trust is to hold predictive and normative expectations about the process and outcome of the interaction. Teng argues that in this, users treat technology analogously to how customers place trust in institutions: "technologies resemble institutions in their design capacity for carrying normative values and inviting relevant expectations about what they are supposed to do" (Teng, 2021). Even though technology is not in itself an institution, it also carries embedded normative values (also shown in the context of trust in engineers, Section "Trust in engineers") and creates expectations about its functionality (as argued in the context of trust in algorithms, Section "Trust in algorithms"). Institutional trust connects trust in structural assurances, i.e. trust in rules and systems, with interpersonal trust (Lahno, 2001). Thus, it allows a more vague, less personal relation between a trustor and a representative; it is enough that the trustee is abstractly characterized, e.g. by their role, to place meaningful normative expectations in them, which in turn invites trusting the technology to φ (Teng, 2023). And further, it also allows placing trust in the technology itself, which invites trust due to its embedded values (Lahno, 2001; Teng, 2021). Hence, treating trust in technology as trust in institutions combines ideas of both conceptions discussed above. In unifying both, it is possible to maintain an indirect interpersonal relationship, warranted by the institutionalized values, but also trusting the artefact directly, while highlighting the embedded normative values. Through the normative expectations towards technology, which are central to institutional trust, the relation is normatively richer than mere reliance. This need for a normatively rich concept is expressed in the exit rules: Criterion (a) demands that a convincing concept of trust is clearly distinguished from reliance. This proposal meets this criterion of a successful conceptual engineering project.

Blockchain applications in particular have attributes that invite expectations: Transparency, resilience, and availability (see Section "Blockchain in a nutshell") promise correct and secure transaction (and storing) of the issued data. In choosing to use a blockchain application, the user, as a trustor, expects the technology to be able to meet these expectations of proper data handling. Further, the engineers and providers of the application are anticipated to have designed it according to those expectations. In this, it may remain pseudonym or unclear (due to the decentralized structure) who the creators and providers are (Teng, 2021, 2023). Blockchain applications incorporate incentives to promote actions benefiting the community, like the gas fee that is paid for successfully mining a new block. As demanded by the institutional account of trust, the users' expectations of the blockchain application are both, predictive (the application will process the data as expected) and normative (the processing and storing of the data follow affirmed values like transparency, equality, or broad availability). Inviting the trustor to have predictive and normative expectations about the technology's functionality, and offering where to find criteria to evaluate it, fits well with the condition of competency (t_ii) to establish trust in technology. However, it is difficult to assess for users, if the algorithms, used in a certain blockchain application, are indeed apt to promote the user's objectives and if the developers followed the perceived values. Thus, the user, as a trustor, still, remains vulnerable in relying on the technology (Jacobs, 2020; Nickel, 2013; Völter et al., 2021). Users lack experience with new technology such as blockchain and often lack the time to

be informed about ever-new developments (Lustig & Nardi, 2015). Thus, the trustor remains vulnerable in using a blockchain application, which means the condition of vulnerability (t_i) is met, too. And in actually using the application also the condition of reliance (t_iii') would be met. Therefore, this proposal does not only fulfil criterion (a) of a promising concept of trust in blockchain applications, i.e. to distinguish trust from reliance (see above), but also it meets criterion (d), to align with the conditions to establish trust. Also, the functions of trust in blockchain applications (cf. criterion (b) for a successful concept) can be explained with this proposal: Trusting the blockchain application allows users with limited understanding of the technology to still use it. They trust the network of nodes and developers to establish the necessary capability and the application in itself to actually φ. This enables these users to act under the risk of using the technology (Teng, 2021). At the same time, it is highlighted, that the organizational structures, including the developers carry responsibility and remain a target of trust (Teng, 2023). Therefore, users are better protected from looking for simplification in the wrong place.

However, due to its complexity, blockchain applications might also mislead normative expectations. Especially the promise of data security and that the blockchain would be free of bias through its decentralized structure is excessive and needs a more differentiated understanding (Hünseler & Pöll, 2023; Teng, 2021). The problem of complexity also hinders the evaluation of the technology's trustworthiness. Users of blockchain applications have a high responsibility to secure their usage. Therefore, some call for support from governmental regulations, certifications, and digested information. Further, they demand engineers and miners to be held more responsible to make blockchain applications (or technology in general) more trustworthy (Nickel, 2013; Tang et al., 2019; Teng, 2021). These demands for evident signs of trustworthiness and embedded values, to which the trustor directs normative expectations, pose a threat to mistakenly assume trust could be induced by trustworthiness; claiming that technology could prompt trust in the trustor, by showing a clear sign of trustworthiness.

Especially in the more technically oriented literature, the impression arises that trust could be created in the trustor by the object of trust; i.e. that trust could be induced by technology. For example, Völter et al. claim that "Signals are an established method to *deliver the trustworthiness* of a technology" (Völter et al., 2021, emphasis added) and Marella et al. even title their paper "Understanding *the creation of trust* in cryptocurrencies: the case of Bitcoin" (Marella et al., 2020, emphasis added). Similar ideas are also put forward by other relevant papers on blockchain applications (Becker & Bodó, 2021; Bundesnetzagentur, 2021; Buterin, 2015). This reverses the direction of the relation in interpersonal trust and comes short of the attributive character of trust. Transforming "A trusts B to φ" into a notion of "B creates trust in A that it will φ". But trust cannot be induced or purposefully created. Thus, it seems necessary to explicitly demand a commitment to highlight the correct direction of the trust-relation, originating from the trustor (cf. criterion (e) to propose a conniving concept of trust in blockchain applications). Trust, as described by conventional accounts, is directed from the trustor to the trustee, placing the trustor explicitly in an active role (Möllering, 2006): The trustor critically evaluates signals from the object of trust and then decides if they indeed place trust that the object will φ or not. This is also implied in the condition of expected performance (t_ii): Based on the signals, the trustor reacts, but the role to determine if the blockchain application is capable to process data properly rests with them. The object of trust can signal trustworthiness, by reminding the trustor of affirming experiences in the past, by promoting embedded normative values, or by certificates proving their competence. But trustworthiness itself cannot be proven or prompted (Reinhardt, 2022; Teng, 2021). At most, the object of trust can invite the trustor to rate it as trustworthy. However, signals of trustworthiness, especially from blockchain applications, are often overestimated in their relevance for trust and thus misleading (Hünseler & Pöll, 2023). And even given that the trustor grants the object trustworthiness—they might still not trust it. Trustworthiness is neither necessary nor sufficient for trust (McLeod, 2021). Hence, it is important to use a clear concept of trust, highlighting the correct direction of the trust-placing-relation.

Trust by normative expectations and attribution

Among the examined proposals, the concept of institutional trust captures best our intuitions towards trust in technology and trust in blockchain applications in particular. Striking is its combination of trust in the technological artefact itself to φ, i.e. in the blockchain application to process and store data correctly and securely, with trust in the organizational structure behind the application. The organizational structure includes software developers, miners and passive node providers, as well as the implemented algorithms and the embedded normative values blockchains promise. Thus, the intuitive notion to trust the blockchain application in itself is captured, while at the same time, providers can be held accountable for their role in creating and maintaining a trustworthy and reliable artefact. The object of trust thus is both, the blockchain application as the technological artefact itself, *and* the structures and individuals responsible for its creation and maintenance. Whether a blockchain application can be considered an institution, as suggested by some accounts, remains unclear. Depending on the research area, different qualities are assumed to be characteristic of institutions (Caporaso & Jupille, 2022). From a broad reading,

blockchain applications could indeed be viewed as institutions: They elicit predictive and normative expectations through built-in qualities (Teng, 2021), and their fundamental design establishes a hierarchical structure within the network of actors, enforced by code, embedding rules, incentives and predefined ends (Ishmaev, 2017; Lahno, 2001). On the others hand however, blockchain applications currently lack the normative influence on society that other definitions require, or the "socially legitimated expertise" institutions should provide (Caporaso & Jupille, 2022). The concept of institution itself appears to require further conceptual engineering. Embracing the understanding of trust in blockchain applications as a form of institutional trust will influence the conditions for developing a convincing concept of institution. Nonetheless, it is sufficient to assume that blockchain applications can be handled analogously to institutions without categorizing them explicitly as such.[12] Thus, a thorough examination of the concept of institution is deferred to future research. To embrace Teng's account[13] of trust in blockchain applications as trust in institutions, the crucial characteristic of the object of trust is to invite the trustor to have normative and predictive expectations. As shown above, blockchain applications fulfil this role.

However, Teng's account does not sufficiently highlight the role of trustor in creating trust. Her examinations in "Towards Trustworthy Blockchains" (Teng, 2021) on how trustworthiness could be established, even seem to fall for the misconception, that trustworthiness might be sufficient to establish trust. In this respect, concerning criterion (e) for a successful concept of trust in blockchain applications, demanding a commitment to the crucial role of the trustor to place trust, her concept proposal should be refined: The advantages of blockchain applications are still often overestimated, and thus it is difficult for users to identify which signals are indeed reliable in inviting trust (Hünseler & Pöll, 2023). Further, the fallacy to curtail the direction of the trust relation as demonstrated above needs to be addressed: The trustor must be placed in an active role, critically evaluating the trust signals blockchain applications might send, and deciding if the application in question indeed seems capable and designed to process the given data adequately and in a second step to rely on the application to do so. Otherwise, "trust would be deterministic and, therefore, a pointless category" (Möllering, 2006). The term "institutional trust" entails a strong conceptual commitment to treat blockchain applications as institutions. As argued above, this commitment is not necessary and makes it necessary to revisit and possibly realign the concept of institution. Therefore, the associative link should be severed and this refined concept proposal renamed. Its most defining aspects are highlighting the role of the trustor to be in the active role of placing trust and that blockchain applications invite normative expectations. Thus, I suggest naming this final concept proposal "Trust by normative expectations and attribution".

The other criteria for a successful concept of trust in blockchain applications are met as well, in a similar manner as by the proposal to treat it as institutional trust: Firstly, trust by normative expectations and attribution can be clearly distinguished from the concept of reliance. Blockchain applications embed values like transparency and thus invite normative expectations. Additionally, due to the complexity and novelty of the technology, users usually remain vulnerable when using blockchain applications, lacking a thorough understanding. Thus, users do not merely have an epistemic relation to blockchain applications, but rather a normatively more significant involvement. Trust in blockchain applications, following this proposal, describes a richer relation than reliance (cf. criterion (a)). Secondly, this proposal sustains the functions of trust in technology (cf. criterion (b)). It explains the advantage for users with limited understanding of the technology of trusting it and remaining able to act despite very limited control over the application's functionality. At the same time, the account of trust by normative expectations and attribution holds the network of nodes and developers responsible for the provision of the functionality. And further, it reminds users to place their trust not only in the code, hoping for trust through security. Thirdly, the account of trust by normative expectations and attribution is explicitly modelled to be applied to blockchain applications, thus fulfilling criterion (c) of a convincing concept proposal for trust in blockchain applications. Fourthly, as shown above, the conditions for establishing trust in technology can be accomplished (cf. criterion (d)). Finally, with the correction compared to the concept proposal of institutional trust, criterion (e) is also achieved. The proposal of trust by normative expectations and attribution highlights the significance of the trustor in granting trust and evaluating the signals the technology may send. It meets the demands for a descriptively and normatively meaningful concept in doing both, capturing the notion of trust that users place in blockchain applications and maintaining a normative meaning of trust beyond mere reliance. Thus, the concept of trust by normative expectation

[12] In her earlier work Teng also conceives „trust in blockchain technology *by analogy* with what we understand of trust in institutions " (Teng, 2021 emphasis added)., even though she later explicitly commits to considering blockchain applications as institutions: "the original blockchain can perform as a virtual institution that users can directly rely upon and interact with" (Teng, 2023).

[13] The institutional account for trust in blockchain systems is mainly put forward by Teng (2021, 2023) and the concept suggested here barely differs from her foundations. However, she did not explicitly consider her suggestion from a conceptual engineering point of view. It therefore seems appropriate to examine her proposal under the criteria mentioned—and to further develop it.

and attribution is a convincing concept proposal for trust in blockchain applications.

Conclusion

This paper set out to adjust the conceptual misalignment of trust in blockchain applications through the methodology of conceptual engineering. The concept of trust was disrupted with the introduction of blockchain applications, by claims that it would render trust obsolete or that it would allow inducing trust automatically. Particularly in the technical literature on blockchain the term "trust" is frequently used, but apparently without necessary normative expressiveness. Trust is described as undesirable, due to its risky nature, and sought to be overcome with the algorithmic corset of blockchain applications. This neglects, however, the necessity of vulnerability of the trustor to establish trust—without which the relation of the user towards the blockchain would be reduced to mere reliance, not trust. Trust cannot be induced or guaranteed from the position of the object of trust. It has to emanate from the trustor, who carefully considers the signals sent by the object, but, after all, autonomously decides to place trust or not. More recent research tendencies claim a shift of trust, rather than its annulation in context of blockchain applications.

The design of this shift should be accompanied by conceptual engineering in order to preserve the normative expressiveness of the concept of trust and to appropriately expand its extension. Conceptual engineering provides tools to investigate the kind of disruption brought about from blockchain applications and to design a convincing solution, by explicating the implications of the concept of trust, the demands for maintaining normative significance and possible side effects of adjusting the concept of trust (cf. Isaac et al., 2022; Löhr, 2023b). The goal is to fix the representational device (i.e. the term "trust") by changing the social norm of how to apply the concept properly and changing relevant normative and empirical expectations (Thomasson, 2021). When expanding the extension of the concept the prescriptive aspect of trust, i.e. how it ought to be used (Isaac et al., 2022), needs to be preserved. This normative demand was expressed as necessary and jointly sufficient conditions to establish trust: Most important is the condition of vulnerability, expressing the necessity that the trustor has limited control over the action they expect to happen, and have to rely on the object of trust to act according to their expectations. The form of the second condition depends on the object of trust: in case the trustor trusts a person directly or a group of people, the second condition describes the expected competence of the trustee(s) to act according to the trustor's expectations. Additional to their competence, the third condition demands that the trustees are also willing to perform the expected action. In case the object of trust is a technological artefact itself, these conditions need to be adjusted, as technological artefacts cannot be willing or competent to do something. Thus, the claim in the second condition is softer: The trustor expects the technology to be capable and designed to act according to the trustor's expectations. In both cases the final condition is of the trustor to indeed rely on the object of trust, binding the previous considerations to an actual action from the trustor. As demonstrated above, not all concept proposals could accomplish these conditions.

Further, some proposals could not preserve the function of the concept of trust in technology, which was described on two levels: On the one hand, trusting technology simplifies the context and allows using technology without necessarily having a thorough understanding of it. By believing that the people behind the technology took care of the functionality or that the artefact in itself is fit to function as expected, the user takes a leap of faith instead having to discern the functioning of the artefact. This instils confidence in the user's action without taking the cognitive effort of deep understanding. On the other hand, trusting technology enables them to act under risk: Users have very limited control of the functionality of the technology they use. So, trust enables them to manage the risk that the technology will not perform as expected and avoid the overwhelm of uncertainty. In context of blockchain applications these functions of trust were threatened by the ambitious promises of proponents, that blockchains could create trust through the security of its algorithms and network. Users might be misled to overlook the necessity to be aware of the risks involved in trusting a technology; that trust is only possible by the trustor taking a vulnerable position towards the application. Further, they might seek to use the application without a deep understanding, neglecting the high responsibility they currently still need to carry. Thus, trust in its function of simplification could cloud the users' awareness of responsibility.

Five criteria were distilled from these considerations, and then applied to three major concept proposals how trust in blockchain applications could be understood, to evaluate their persuasiveness each. First, trust in engineers was inspected. Offering an account of interpersonal trust in the people behind the blockchain (miners, software engineers, node providers, etc.). This concept should be rejected as the actual willingness of the trustee to φ cannot be evaluated. In the second concept proposal, trust was not placed in the technology qua the people behind it, but directly in the artefact. Trust in algorithms could account for the intent to not merely place trust in the *creation* of functionality but in the actual functioning of the technology. This view invited the understanding to use security as a promising way to manage risks in transactions, but to establish trust through security is bound to fail: Eliminating the vulnerability of the trustor

leads to the loss of trust, not the creation of it. Vulnerability in relying on technology is a necessary condition to place trust. Furthermore, supporters of "in proof we trust" in particular seem to succumb to the false impression that trust can be induced in the trustor by technology. Hence, also this concept must be rejected. The third proposal provides a combination of the advantages of both other conceptions: blockchain applications are trusted by placing trust in the system and its code on the one hand, and in the network of people behind the application on the other hand. Further, they trust the institutional character of the application in guiding the interplay of both these aspects by embedded normative values. However, it remains unclear if blockchain applications can indeed be handled as institutions or merely analogously. Further, the concept of institutional trust does not sufficiently address the fallacy to curtail the direction of trust. Therefore, a fourth proposal for the concept of trust in blockchain applications was put forward, which draws from the concept of institutional trust, but emphasizes the active role of the trustor. In taking this role seriously, the concept of trust by normative expectations and attribution highlights that technology and blockchain applications can signal trustworthiness, but it remains up to the trustor to evaluate those and grant trust. This revised concept should help to clarify some ambiguities and offer a more meaningful use of the term "trust".

The aim of this text was twofold: Firstly, it intended to provide a proposal of an adjusted concept of trust in the context of blockchain applications. Secondly, it is meant to convince members of the discussion around blockchain applications of the relevance of a revised use to the term "trust", highlighting the unintended implications otherwise (cf. Isaac et al., 2022). To implement this adapted concept I hope to convince researchers especially in computer science of a more constrained use of the word and invite ethicists to highlight the implications for technology, so that gradually normative and empirical expectations from the concept of trust in blockchain applications are adjusted (cf. Thomasson, 2021).

Still, blockchain applications remain to be a very complex technology. As blockchain proponents give high promises, that they cannot all keep, users are sent mixed signals of trustworthiness. Additionally, to the difficulty of establishing a sufficient understanding of this complex technology, user might be rendered incapable to form the expectations that are demanded by the concept. They might not actually be able to evaluate if an application is sufficiently likely to perform as they expect and if it is the best option to achieve their goal; it is not even certain if they can form accurate expectations in the first place. Thus, further empirical research is necessary to ground the assumption in data. Based on these results, the suggested concept of trust in blockchain applications by normative expectation and attribution should be revisited and investigated if the proclaimed forming of relevant expectations is possible. If not, the concept of trust in blockchains, as presented here, needs to be rejected and revised. Also, it should be looked into the concept of institution. It would be interesting if either blockchain technology or blockchain applications could be constructed as institution and under which conditions. If this is possible, it would be worthwhile to model trust in blockchain applications more closely as institutional trust and examine the implications. This would certainly provide further insights into the concept of trust in blockchain applications and further sharpen the term.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Alfano, M., & Huijts, N. (2022). Trust and distrust in institutions and governance. In J. Simon (Ed.), *Handbook of trust and philosophy*. Routledge.

Baier, A. (1986). Trust and antitrust. *Ethics, 96*(2), 231–260.

Becker, M., & Bodó, B. (2021). Trust in blockchain-based systems. *Internet Policy Review, 10*(2), 1–10. https://doi.org/10.14763/2021.2.1555

Bundesnetzagentur. (2021). *Die Blockchain-Technologie. Grundlagen, Potenziale und Herausforderungen*. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen. https://www.bundesnetzagentur.de/DE/Fachthemen/Digitalisierung/Technologien/Blockchain/Links_Dokumente/einfuehrung_bc.pdf?__blob=publicationFile&v=12

Buterin, V. (2015). Visions, part 2: The problem of trust. *Ethereum foundation blog*. https://blog.ethereum.org/2015/04/27/visions-part-2-the-problem-of-trust

Caporaso, J. A., & Jupille, J. (2022). Definitions of institutions. In *Theories of institutions* (pp. 159–164). Cambridge University Press. https://doi.org/10.1017/9781139034142.007

Carter, J. A., & Simion, M. (2020). The ethics and epistemology of trust. In J. Fieser & B. Dowden (Eds.), *Internet encyclopedia of philosophy*.

Finck, M. (2018). *Blockchain regulation and governance in Europe* (1st ed.). Cambridge University Press.

Friedman, B., Khan, P. H., Jr., & Howe, D. C. (2000). Trust online. *Communications of the ACM, 43*(12), 34–40. https://doi.org/10.1145/355112.355120

Fries, I. (2022). »In Code We Trust«? *Zeitschrift für Evangelische Ethik*, *66*(4), 264–276. https://doi.org/10.14315/zee-2022-660405

Frizzo-Barker, J., Chow-White, P. A., Adams, P. R., Mentanko, J., Ha, D., & Green, S. (2020). Blockchain as a disruptive technology for

business: A systematic review. *International Journal of Information Management, 51*, 102029. https://doi.org/10.1016/j.ijinfomgt.2019.10.014

Hawley, K. (2014). Trust, distrust and commitment. *Noûs, 48*(1), 1–20. https://doi.org/10.1111/nous.12000

Himmelreich, J., & Köhler, S. (2022). Responsible AI through conceptual engineering. *Philosophy & Technology*. https://doi.org/10.1007/s13347-022-00542-2

Hoffman, R. (2014, November 17). *The future of the bitcoin ecosystem and 'Trustless Trust'—Why I invested in blockstream*. https://www.linkedin.com/pulse/20141117154558-1213-the-future-of-the-bitcoin-ecosystem-and-trustless-trust-why-i-invested-in-blockstream

Hopster, J., & Löhr, G. (2023). Conceptual engineering and philosophy of technology: Amelioration or adaptation? *Philosophy & Technology, 36*(4), 70. https://doi.org/10.1007/s13347-023-00670-3

Hünseler, M., & Pöll, E. (2023). Promises and problems in the adoption of self-sovereign identity management from a consumer perspective. In F. Bieker, J. Meyer, S. Pape, I. Schiering, & A. Weich (Eds.), *Privacy and identity management* (pp. 85–100). Springer. https://doi.org/10.1007/978-3-031-31971-6_8

Hyrynsalmi, S., Hyrynsalmi, S. M., & Kimppa, K. K. (2020). *Blockchain ethics: A systematic literature review of blockchain research*. Springer.

Isaac, M. G., Koch, S., & Nefdt, R. (2022). Conceptual engineering: A road map to practice. *Philosophy Compass, 17*(10), 1–15. https://doi.org/10.1111/phc3.12879

Ishmaev, G. (2017). Blockchain technology as an institution of property. *Metaphilosophy, 48*(5), 666–686. https://doi.org/10.1111/meta.12277

Jacobs, M. (2020). How implicit assumptions on the nature of trust shape the understanding of the blockchain technology. *Philosophy and Technology, 34*(3), 573–587. https://doi.org/10.1007/s13347-020-00410-x

Kelly, H., & Guskin, E. (2021). Americans widely distrust Facebook, TikTok and Instagram with their data, poll finds. *Washington Post*. https://www.washingtonpost.com/technology/2021/12/22/tech-trust-survey/

Kelp, C., & Simion, M. (2023). What Is trustworthiness? *Noûs, 00*, 1–17. https://doi.org/10.1111/nous.12448

Koch, S., Löhr, G., & Pinder, M. (2023). Recent work in the theory of conceptual engineering. *Analysis, 83*(3), 589–603. https://doi.org/10.1093/analys/anad032

Lahno, B. (2001). Institutional trust: A less demanding form of trust? *Revista Latinoamericana De Estudios Avanzados RELEA, 15*, 19–58.

Lipman, M. A. (2023). On bitcoin: A study in applied metaphysics. *The Philosophical Quarterly*. https://doi.org/10.1093/pq/pqad030

Löhr, G. (2023a). Conceptual disruption and 21st century technologies: A framework. *Technology in Society, 74*, 102327. https://doi.org/10.1016/j.techsoc.2023.102327

Löhr, G. (2023b). If conceptual engineering is a new method in the ethics of AI, what method is it exactly? *AI and Ethics*. https://doi.org/10.1007/s43681-023-00295-4

Luhmann, N. (1968). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*. Enke.

Lustig, C., & Nardi, B. (2015). Algorithmic authority: The case of bitcoin. *2015 48th Hawaii International Conference on System Sciences*, 743–752. https://doi.org/10.1109/HICSS.2015.95

Mallard, A., Méadel, C., & Musiani, F. (2014). *The paradoxes of distributed trust: Peer-to-peer architecture and user confidence in bitcoin*. http://peerproduction.net/issues/issue-4-value-and-currency/peer-reviewed-articles/the-paradoxes-of-distributed-trust/

Marella, V., Upreti, B., Merikivi, J., & Tuunainen, V. K. (2020). Understanding the creation of trust in cryptocurrencies: The case of bitcoin. *Electronic Markets, 30*(2), 259–271. https://doi.org/10.1007/s12525-019-00392-5

McKnight, D., Carter, M., Thatcher, J., & Clay, P. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems, 2*, 12–32. https://doi.org/10.1145/1985347.1985353

McLeod, C. (2021). Trust. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2021)*. Stanford University.

Möllering, G. (2006). Trust, institutions, agency: Towards a neoinstitutional theory of trust. In R. Bachmann (Ed.), *Handbook of trust research*. Edward Elgar Publishing.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system* [Whitepaper]. https://bitcoin.org/bitcoin.pdf

Nickel, P. J. (2013). Trust in technological systems. In M. J. de Vries, S. O. Hansson, & A. W. M. Meijers (Eds.), *Philosophy of engineering and technology* (pp. 223–237). Springer.

Nickel, P. J. (2021). Trust in engineering. In D. Michelfelder & N. Doorn (Eds.), *Routledge handbook of the philosophy of engineering* (pp. 494–505). Routledge.

Nickel, P. J., Franssen, M., & Kroes, P. (2010). Can we make sense of the notion of trustworthy technology? *Knowledge, Technology & Policy, 23*(3), 429–444. https://doi.org/10.1007/s12130-010-9124-6

Nissenbaum, H. (1999). Can trust be secured online? A theoretical perspective. *Etica E Politica*, *1*(2).

O'Neill, O. (2002). *A question of trust: The Bbc reith lectures 2002*. Cambridge University Press.

O'Neill, O. (2020). *Questioning trust*. Routledge.

Ostern, N. (2018). *Do you trust a trust-free technology? Toward a trust framework model for blockchain technology completed research paper*. 39. International Conference on Information Systems (ICIS), San Francisco.

Pesch, P. J. (2019). Blockchain, smart contracts und Datenschutz. In M. Fries & B. P. Paal (Eds.), *Smart contracts* (pp. 13–24). Mohr Siebeck.

Popper, N. (2021). *Lost passwords lock millionaires out of their bitcoin fortunes*. The New York Times.

Reinhardt, K. (2022). Trust and trustworthiness in AI ethics. *AI and Ethics*. https://doi.org/10.1007/s43681-022-00200-5

Schlatt, V., Schweizer, A., Urbach, N., & Fridgen, G. (2016). *Blockchain: Grundlagen, Anwendungen und Potenziale* [White Paper]. Projektgruppe Wirtschaftsinformatik des Fraunhofer-Instituts für Angewandte Informationstechnik FIT. https://publica.fraunhofer.de/handle/publica/298479

Smits, M., & Hulstijn, J. (2020). Blockchain applications and institutional trust. *Frontiers in Blockchain, 3*, 5. https://doi.org/10.3389/fbloc.2020.00005

Tallant, J. (2019). You can trust the ladder but you shouldn't. *Theoria, 85*(2), 102–118. https://doi.org/10.1111/theo.12177

Tang, Y., Xiong, J., Becerril-Arreola, R., & Iyer, L. (2019). Ethics of blockchain: A framework of technology, applications, impacts, and research directions. *Information Technology & People, 33*(2), 602–632. https://doi.org/10.1108/ITP-10-2018-0491

Teng, Y. (2021). Towards trustworthy blockchains: Normative reflections on blockchain-enabled virtual institutions. *Ethics and Information Technology, 23*(3), 385–397. https://doi.org/10.1007/s10676-021-09581-3

Teng, Y. (2023). What does it mean to trust blockchain technology? *Metaphilosophy, 54*(1), 145–160. https://doi.org/10.1111/meta.12596

The Trust Machine. (2015). *The economist*. https://www.economist.com/leaders/2015/10/31/the-trust-machine

Thomasson, A. (2021). Conceptual engineering: When do we need it? How can we do it? *Inquiry*. https://doi.org/10.1080/0020174X.2021.2000118

Townley, C., & Garfield, J. L. (2013). Public trust. In P. Mäkelä & C. Townley (Eds.), *Trust: Analytic and applied perspectives* (pp. 95–107). Brill.

Völter, F., Urbach, N., & Padget, J. (2021). Trusting the trust machine: Evaluating trust signals of blockchain applications. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2021.102429

Walch, A. (2019). In code(rs) we trust: Software developers as fiduciaries in public blockchains. In P. Hacker, I. Lianos, G. Dimitropoulos, & S. Eich (Eds.), *Regulating blockchain: Techno-social and legal challenges.* Oxford University Press.

Wang, W., Lumineau, F., & Schilke, O. (2022). Blockchains: Strategic implications for contracting, trust, and organizational design. *Cambridge University Press*. https://doi.org/10.1017/9781009057707

Werbach, K. (2018). *The blockchain and the new architecture of trust*. The MIT Press.

Wingreen, S. C., & Baglione, S. L. (2005). Untangling the antecedents and covariates of e-commerce trust: Institutional trust vs knowledge-based trust. *Electronic Markets, 15*(3), 246–260. https://doi.org/10.1080/10196780500209010

Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). *Blockchain technology overview (Internal Report NIST IR 8202)*. National Institute of Standards and Technology, U.S. Department of Commerce.