# On the Establishment of Trust: Challenges, Opportunities and Socio-Cultural Factors

**Christoph Lipps[1], Denise Scharwatz[2] and Henry Collier[3]**
[1]Intelligent Networks Research Group, German Research Center for Artificial Intelligence, Kaiserslautern, Germany
[2] RPTU University Kaiserslautern-Landau, Kaiserslautern, Germany
[3] Thomas Edison State University, Trenton, USA

Christoph.Lipps@dfki.de
D.Scharwatz@rptu.de
hcollier@tesu.edu

**Abstract:** Trust is one of the fundamental necessities of human beings and, according to *Stephen R. Covey*, not simply the "glue of life", but also the "most essential ingredient in effective communication". Even though the principles and importance of trust are as old as humanity itself -in ancient times, trusting strangers could mean the difference between life and death, and thus pose an immediate threat to one's social tribe- trust is gaining increasing attention, particularly in light of tomorrow's all-electric society and the decentralization and globalization associated with it. Contrary to previous decades, physical proximity is no longer necessary to access a system; access is possible (almost) anytime, (almost) anywhere. However, trust is a multidimensional concept depending on a multitude of aspects, such as the specific application, the value of the resource, and the available technology, but also -for instance- on the people who are managing access to systems and applications. As societies become increasingly aware of data breaches, algorithmic surveillance, and commercialization of personal data, cultural values have shifted toward individual agency and distrust of centralized institutions. As trust could never be taken for granted it rather must be earned, orchestrated and continuously verified. Technologies like multi-factor authentication (MFA) reflect this, offering multi-layered mechanisms for identity verification and access protection. These developments illustrate a broader societal reshaping of trust -from implicit trust toward conditional, data-driven security- and show how technological design is shaped by evolving social anxieties and expectations. Against this background, this work focuses on socio-cultural influences on the definition of trust and trustworthiness, such as origin (geography, culture, political system) or educational background and training. In particular, it follows research questions including: i) How do socio-cultural perceptions of trust and identity shape the development and adoption of digital security technologies such as MFA; ii) in what/which ways does the rise of digital surveillance and data breaches influence societal expectations of privacy and trust in technological systems?; and iii) how is the concept of trust redefined in the digital age, and what role do authentication technologies play in mediating this transformation?

**Keywords:** Trust**,** Trustworthiness, Critical infrastructure security, Cyber security and resilience

## 1. Why Trust Matters: Geopolitical Drivers for Trust and Trustworthiness in Cybersecurity

The conceptual foundations of trust in digital systems are undergoing profound transformation. While *trust* traditionally refers to a subjective attitude toward people, systems, or institutions, *trustworthiness* denotes the objective characteristics and behaviors that render an entity reliable (Fettweis, et al., 2025). This distinction has become vital in a digital society where interactions are increasingly mediated by complex technical infrastructures rather than direct human relationships. Trustworthiness must therefore be established, demonstrated, and continuously verified. Security mechanisms such as multi-factor authentication (MFA) illustrate this evolution: they do not create trust themselves, but provide measurable, verifiable properties -such as strong identity assurance- strengthening the perception of trustworthiness. As digital infrastructures become more interconnected and exposed to sophisticated threats, the capability to design and evaluate trustworthiness has emerged as a central prerequisite for secure communication and resilient critical infrastructures (Sithungu & Lipps, 2025).

Building on this distinction, trust can be understood -following (Luhmann, 2017)- as a subjective expectation that enables individuals to navigate complexity by relying on assumptions about the behavior of others or of systems. It is shaped by prior experience, perceived risk, and contextual interpretation. Trustworthiness, by contrast, refers to the objectively assessable qualities of a system or entity, expressed through measurable attributes such as reliability, security assurances, or compliance with standards like those defined by National Institute of Standards and Technology (NIST) -e.g. SP 800-160, SP 800-53, SP 800-12-, and International Organization for Standardization (ISO), -e.g. ISO/IEC TS 5723:2022. Although these two concepts are often implicitly treated as mutually reinforcing, their relationship is neither linear nor guaranteed. A system that is demonstrably more trustworthy in a technical sense does not necessarily evoke higher levels of human trust, particularly when cognitive biases, sociocultural factors, or limited transparency hinder users' ability to interpret

or appreciate such assurances. This conceptual tension is increasingly relevant in digital environments where subjective trust judgments depend heavily on abstract indicators of technical trustworthiness rather than direct interpersonal experience.

Against this backdrop, trust and trustworthiness represent two of the most strategically important -yet increasingly fragile- pillars of modern cybersecurity. Rising geopolitical tensions and accelerating digital interdependence challenge the ability of states, organizations, and operators of critical infrastructure to build and maintain trustworthy technological ecosystems. Recent incidents illustrate the erosion of this trustworthiness: the compromise of a Norwegian dam's control system in April 2025 through weak access credentials (Pritchard, 2026); the "Salt Typhoon" attacks on U.S. telecommunications providers, which targeted identity verification and secure communication systems (Boston Institute of Analytics, 2025); and the ransomware attack on Colonial Pipeline in 2021, one of the most impactful cyberattacks on U.S. energy infrastructure (Lipps et al., 2022). These cases demonstrate how even seemingly isolated intrusions can reveal systemic weaknesses, marking a broader trend in which cyber operations increasingly target the integrity, availability, and reliability of essential services rather than merely extracting data.

At the same time, the global threat landscape is shifting rapidly. The rise of generative AI (GenAI) has changed the nature of cyberattacks: attackers now automate intrusion attempts, craft convincing social engineering campaigns, and conduct long-term operations designed to undermine institutional trust rather than simply compromise systems. Identity-based intrusions exploiting valid credentials, trusted AI agents, or federated identity platforms are proliferating, eroding traditional assumptions about trust boundaries and exposing the limitations of perimeter-based security models.

Critical infrastructures -energy grids, water treatment facilities, transportation networks, and healthcare systems- are particularly vulnerable. Many rely on outdated operational technologies (OT) never designed for secure connectivity. Regulatory responses such as the NIS2 Directive (NIS 2 Directive, 2022) and the Cyber Resilience Act (Cyber Resilience Act, 2024) reflect the urgency of strengthening systemic trustworthiness, but compliance alone cannot guarantee resilience. Instead, deeper insight is required into how trust and trustworthiness can be established and maintained across distributed, heterogeneous, and geopolitically sensitive environments.

Simultaneously, emerging technologies create new trust dependencies. AI-enabled security systems offer improved detection capabilities but introduce opaque decision-making processes and new attack surfaces (Chokkanathan et al., 2024). Research on AI-driven Zero Trust architectures highlights the necessity of continuous verification, identity assurance, and adaptive trust assessment—principles especially relevant as adversaries become more dynamic. Likewise, the global transition toward postquantum cryptography reflects concerns that quantum capable adversaries could compromise today's cryptographic foundations.

The interplay of these developments -geopolitical fragmentation, AI-assisted cyber operations, vulnerable infrastructure, and the shift toward quantum-resistant security-, creates an urgent need for updated frameworks to strengthen trustworthiness. Such frameworks must be resilient to deception, interoperable across jurisdictions, and capable of functioning even when traditional trust anchors, such as state cooperation or vendor reliability, are destabilized. Building trustworthiness is therefore not merely a technical task, but a strategic requirement for societal stability, economic continuity, and national security.

Accordingly, this work examines how trust and trustworthiness can be systematically conceptualized, built, and maintained in modern communications infrastructures and critical systems. Integrating insights from cybersecurity engineering, policy studies, and emerging-technology research, it aims to provide a comprehensive foundation for designing trust mechanisms capable of withstanding the pressures of contemporary geopolitical and technological realities.

## 2. Definition of Trust and Trustworthiness

Trust and trustworthiness are often conflated, yet they refer to fundamentally different concepts. Thus, there are very diverse perspectives on the concept of trust, as it

- is a belief, or more importantly a psychological state where someone believes in the integrity, reliability or ability of someone else (TrustTalk, 2026),
- results in one person's willingness to rely upon the other person,
- is a foundational component of personal, business, and societal relationships, to include the relationships within oneself, and

- is a crucial component of life, especially one's social life (Yamamoto, 1990).

So, trust has several components that are interconnected. These components include reliability, honesty, competence, benevolence and integrity. As humans, we are taught to trust other people, to trust situations and to trust that good will prevail over evil. However, from a cybersecurity perspective, trust is most elusive, and misguided trust leads to security incidents.

Trust can be further categorized into three categories: i) logical trust, ii) emotional trust and iii) relationship trust. Logical trust and emotional trust are considered individual trust attributes, while relationship trust is its own category called relational trust attribute (Cho, et al., 2015). Whereas trustworthiness is a quality or characteristic that makes someone worthy of trust. Trust is something that is built based on information from different sources to include a person's reputation, past behaviour, identity, communication and social norms. For example, a religious leader is typically considered to be someone who can be trusted, this is an example of past behaviour, identity and social norms. As a result, a religious leader is considered trustworthy. Trustworthiness is a primary factor in developing and maintaining trust and trusting. Trust and trustworthiness are influential factors impacting effective communication, cooperation, and decision-making (TrustTalk, 2026).

Who we trust and why we trust depends on what kind of person we are as an individual. There is no cookie cutter concept that applies to everyone. At its core, trust is unequivocably influenced by culture. Culture is one of the strongest influences on decision-making. Culture comes in many forms, two of the primary forms are individualism and collectivism. In both forms of culture, the decision-making process is influenced differently. In individualism, the cultural influences tend to lead people to make decisions that result in personal benefit while in the collectivism form, the result is decisions that group benefit. Within each of these forms, there are many sub forms of culture as well, resulting in varied outcomes from within one primary form of culture (Glazer & Karpati, 2014) (Yates & De Oliveira, 2016). A clear example of how culture influences trust and trustworthiness can be found in how people use cultural experiences to frame an individual's understanding of people and situations. Since trust is a psychological state of mind and trustworthiness is quality that makes someone worthy of trust, both are based on a person's experience with someone or a situation (Glazer & Karpati, 2014).

In the world of information technology and cybersecurity, trust and trustworthiness are constantly being evaluated and critiqued. In modern cybersecurity, the community increasingly adopts what is known as a *zero-trust* perspective. While this approach is often described -somewhat misleadingly- as trusting "only what can be verified," it is important to clarify that absolute verification is neither possible nor part of the zero-trust paradigm. The term "zero trust" is a marketing shorthand rather than a literal description: the underlying mechanisms do not eliminate trust but rather aim to minimize the trust base and make trust assumptions explicit, for example by shifting reliance from hardware trust anchors to mathematically verifiable properties. Nonetheless, a residual set of unverified assumptions remains unavoidable, and their extent is not always fully knowable.

This becomes clear when considering practical scenarios. For instance, a recruiter at a leading company must process applications from individuals they have never met. Their role requires opening résumés or portfolios in formats such as PDF or Word -files that inherently carry some risk because they may contain malicious content (Naprys, 2025). Zero-trust concepts can help reduce exposure, but they cannot eliminate underlying trust dependencies entirely. The aim of this work is not to propose a solution to this scenario but to highlight how such challenges reveal ongoing tensions in cybersecurity, especially concerning how the community conceptualizes trust and trustworthiness.

Even in a zero-trust environment, breaches remain possible -not only through human actions, but more generally through the breakdown or violation of underlying trust assumptions, many of which may be implicit or unknown. Social engineering illustrates this challenge: it manipulates individuals into taking actions they otherwise would not, thereby exploiting these residual trust dependencies. Social engineers use all the tools in the book, to include manipulation and deception resulting in attacks that use a person's behaviours and emotions against them to succeed. Social engineers certainly use trust as a weapon, and they are very successful at it. Trust, curiosity, fear and urgency are key components in social engineering techniques that are successful (Pujari & Hussain, 2024). How do social engineers develop trust with a victim, resulting in them being trustworthy in the eyes of the victim. First and foremost, they learn what motivates the person. They then use this information to develop a convincing scam that pushes the person's buttons, using fear and urgency as primary factors.

People become victims of social engineering because their decision-making process has been manipulated. As already mentioned, trust is a key component of the decision-making process. If you trust someone, or something, you are more likely to click on a link or provide your user credentials. This is why many social engineering attacks appear to be people of authority like the Chief Executive Officer (CEO) or the IT department. We are taught to inherently trust these individuals because their position makes them trustworthy and humans are often lazy and do not verify that the email is actually from the person it says it is from.

Trust and trustworthiness are complicated and have been studied many times in various fields (Cho, et al., 2015). All three of the previously noted categories of trust (logical, emotional and relationship) are used by social engineers to conduct their nefarious activities and social engineers implement this attack on trust without care, concern or compassion. Their only goal is to successfully obtain what they want, usually money.

People being vulnerable to social engineer attacks, like phishing, is the direct result of cognitive biases in the decision-making process, for example, inherently trusting that people are good (Pujari & Hussain, 2024). These biases are the result of many social and cultural influences, including religion and social engagement that directly impact trust and trustworthiness.

These dynamics illustrate that trust and trustworthiness operate simultaneously at the human, cultural, and organizational levels. To understand how these concepts manifest in modern digital systems, it is necessary to examine their technological foundations as well as the broader socio-economic forces that shape how trust is created, maintained, and challenged in contemporary society.

## 3. Technological Foundations of Trust and Trustworthiness

But, things are somewhat different in digital and technical ecosystems, where trustworthiness is not an abstract concept, but a feature established and determined by verifiable and quantifiable mechanisms. Whereas, as mentioned above, human trust is primarily based on perception and experience, technological trustworthiness depends on objective characteristics that can be measured, validated, continuously revalidated, and monitored.

These aspects can be divided into four major components, which are discussed in more detail below and which operationalize different facets of trust and trustworthiness in modern infrastructures: i) the conceptual foundations of technological trust, ii) identity and access management, iii) cryptographic security, and iv) trusted execution environments (TEEs). Whereas it should be emphasized that the term “trusted” in TEE does not imply inherent trustworthiness. In the terminology of trusted computing, TEEs belong to the Trusted Computing Base (TCB), components that must be trusted because they are capable of violating the security policy, not because their behavior can be fully verified. TEEs thus exemplify the unavoidable trust assumptions that persist even in systems designed to maximize verifiable trustworthiness (Lipps, et al. 2018).

At its core, technological trust relies on a set of security and protection goals that ensure systems behave reliably and predictably. While confidentiality, integrity, and availability (CIA) (Schneier, 1995) form a foundational subset -ensuring that data remains private, unaltered, and accessible when needed- these represent only part of the broader landscape. Depending on the specific use case, additional properties such as accountability, authenticity, anonymity, unlinkability, and non-repudiation may be equally essential, as trust cannot be established without the relevant combination of these conditions. However, trustworthiness extends beyond these fundamentals. It encompasses transparency, predictability, and resilience under stress (Van der Ham, 2021). Systems have to prove that they actually deliver defined and warranted performance, are resistant to manipulation (and can detect attempts at manipulation), and can be recovered after failures. This shift from implicit to explicit, evidence-based trust is crucial in an age where interactions take place via complex, distributed infrastructures rather than direct human relationships. Trustworthiness thus becomes a function of verifiable characteristics: secure configurations, verifiable processes, and continuous compliance with defined security principles.

Another essential element is *identity assurance* -that is, the reliable binding of actions to an authenticated principal when accountability is required- because, without it, many security controls become fragile; however, identity in this sense does not preclude anonymity or pseudonymity, which can provide strong security guarantees in use cases (e.g., voting) where linkability to a real-world person is neither necessary nor desirable (Finney & Kindervag, 2023).

Modern identity and access management (IAM) systems implement MFA schemes to reduce reliance on single sign-on credentials by combining knowledge factors (passwords), possession factors (tokens), and inherence factors (biometrics) (Lipps, Herbst, and Schotten, 2021). Beyond MFA, adaptive authentication evaluates

contextual signals -device status, geolocation, behavioral patterns- to dynamically adjust the level of trust. Federated identity systems (e.g., Security Assertion Markup Language (SAML), OAuth 2.0, OpenID Connect) enable cross-domain interoperability but introduce new trust dependencies that require robust governance and cryptographic safeguards for token issuance and revocation. Besides, privileged access management (PAM) adds another layer by controlling high-risk accounts through vaulting, session recording, and just-in-time elevation. Furthermore, Zero Trust enforces continuous authentication and continuous authorization, which are related but not identical. A request is re-authenticated using up-to-date identity and posture signals, and separately re-authorized against policy and risk -even in designs that permit authorization without revealing a civil identity (e.g., pseudonymous credentials). Identity signals become dynamic inputs to authorization, so trust is continually re-evaluated rather than granted by static entitlements (Gambo & Almulhem, 2025).

The technological backbone of this trust lies in the use of cryptographic methods, which provide *conditional* mathematical assurances based on widely accepted computational hardness assumptions. While only very few cryptographic constructs offer formally proven guarantees, modern cryptography enables strong practical protections: encryption schemes preserve confidentiality under assumed adversary limits, and digital signatures or message authentication codes (MACs) provide assurance of integrity and origin. In this sense, cryptography does not eliminate uncertainty, but it offers rigorously analyzed mechanisms that reduce reliance on blind trust by grounding security properties in well-understood mathematical assumptions. Public key infrastructure (PKI) extends these safeguards to distributed systems, enabling secure key exchange and certificate-based chains of trust. However, cryptographic trust is only as strong as its key management practices: secure generation, storage, rotation, and revocation are essential to prevent silent compromises. New threats from quantum computers challenge traditional algorithms such as RSA and ECC, requiring the development of post-quantum cryptography (PQC). Hybrid cryptographic methods that pair classical and post-quantum algorithms are increasingly recommended -not because they ensure forward security, but because they reduce the risk of regression during the transition. Such hybrid constructions require an adversary to compromise both the classical and PQC algorithms, thereby avoiding a scenario where reliance on a single primitive (classical or PQC) could be invalidated by unforeseen cryptanalytic advances. These measures demonstrate that cryptographic trust is not static but must evolve with the capabilities of attackers and be continuously strengthened (Lipps et al. 2020).

While cryptography safeguards data during transmission and in storage, Trusted Execution Environments protect data during use (Shepherd, C., and Markantonakis, 2024). TEEs create isolated enclaves within processors where sensitive computations take place, shielded from the operating system and other applications. This isolation prevents attackers -even those with root access- from manipulating or observing critical processes. TEEs also enable remote verification, allowing systems to prove to external auditors that they are executing approved code in a secure state. This capability is critical for distributed infrastructures where trust must exist across organizational and geographic boundaries. In cloud and edge computing, TEEs support confidential computing and ensure that workloads remain trustworthy even in multi-tenant environments. By anchoring trust in the hardware, TEEs provide a foundation for higher levels of security and complement cryptographic and identity-based controls.

## 4. The Socio-economic Aspects of Trust and Trustworthiness

Trust and trustworthiness in digital systems emerge not only from technical assurances but also from socio-economic contexts that shape how users perceive, interpret, and act upon those assurances. While engineers ground trustworthiness in measurable properties -as discussed above- users rely on heuristics influenced by cultural norms, education, institutional credibility, and experiences with misinformation. This creates a gap between *engineered trustworthiness* and *perceived trust*, making it essential to understand the socio-economic factors that mediate both.

Empirical evidence shows how strongly social interpretation influences trust. A mixed-methods study conducted in the United Kingdom with 451 respondents demonstrates that many individuals attribute higher trust to AI systems than to humans, largely because AI is perceived as less emotional and less susceptible to manipulation (Gerlich, 2024). This illustrates how users may interpret algorithmic systems through socio-emotional expectations rather than technical guarantees. Conversely, negative experiences can rapidly erode trust. Furthermore, a major U.S. study of 1,045 teens and parents highlights that exposure to deepfakes and fabricated content reduces confidence in information quality, often prompting broader distrust in digital platforms and technologies because users begin doubting the authenticity and provenance of online material (Baron et al., 2025). In both cases, perceptions -not technical properties- drive trust outcomes.

Besides, education and digital literacy shape how people assess trust in uncertain situations. Users with higher media literacy are more capable of distinguishing authentic content from manipulated content, reducing both naïve over trust and unwarranted skepticism. However, most adolescents struggle to determine when content is AI generated, as shown in the same U.S. study material (Baron et al., 2025). This suggests that without adequate training, users face an increasing verification burden they cannot manage alone, especially as generative AI accelerates content production.

Cultural and political contexts further moderate trust: Social norms may emphasize personal autonomy, communal values, or institutional authority, all of which influence how users interpret technological trustworthiness. For example, individuals embedded in environments that value impartiality may more readily accept AI systems as "neutral," whereas individuals in contexts characterized by skepticism toward centralized institutions may distrust AI due to broader political narratives. These orientations help explain why trust levels differ sharply across countries and demographic groups.

Direct interaction with AI systems is another affecting factor of trust, as research shows that users do not necessary require sth. like chatbots to behave human-like and thereby establish trust, but instead, creating interfaces that convey *social presence* -that is, a sense of being emotionally responsive and socially attuned- is more effective than anthropomorphic mimicry (Huynh & Aichner, 2025). This nuance is vital: trust is shaped by *interactional cues*, not by surface-level human imitation. Social presence reduces uncertainty by making AI feel embedded in a familiar relational context, thus reinforcing perceived trustworthiness.

Sector-specific dynamics highlight further socio-economic factors. In healthcare, scenario-based experiments with medical students and physicians demonstrate that explainability and deep integration of AI into clinical workflows significantly increase trust in AI-based clinical decision support systems (AI-CDSS). However, these same features may also heighten professional identity threats, as clinicians worry about losing autonomy or expertise (Ackerhans et al., 2025). Interestingly, the study shows that trust in the system can partially mitigate identity threat, yet accountability mechanisms -such as requiring physician signatures on AI-generated recommendations- can intensify it (Ackerhans et al., 2025). This reveals a complex interplay between trust, responsibility, and professional role identity.

Institutional trust is equally influential at a societal level. A U.S. survey examining public attitudes toward AI governance finds that trust in AI companies correlates with reduced support for slowing or regulating AI development, suggesting that some individuals believe industry self-regulation is sufficient when companies are considered trustworthy (Bullock et al., 2025). Conversely, individuals who trust government institutions more strongly tend to support stricter regulation. These findings show that public policy preferences around AI are not driven solely by risk assessments but also by trust in *the actors* responsible for managing those risks.

Taken together, these findings demonstrate that trust and trustworthiness in digital systems cannot be understood in technical terms alone. They are also shaped by cultural expectations, individual capabilities, professional identities, and institutional narratives. Engineering trustworthy systems remains essential -but achieving *societal* trust additionally requires transparent governance, education, and design approaches that account for human expectations and cognitive tendencies. Ultimately, trust is not simply a security property to be built; it is a socio-technical achievement that must be continually earned.

## 5. Use-Cases

Trust in digital systems and applications is, as mentioned, a multidimensional concept. It is therefore not sufficient to implement advanced cryptographic protocols and enforce zero-trust principles; trust must also be embedded in organizational processes, cultural norms, and human behavior. By describing the discussion using two application scenarios, the main objectives are: i) to illustrate theoretical concepts using concrete examples; ii) to highlight practical challenges that arise when technology encounters sociocultural diversity; and iii) to provide guidelines for designing trust frameworks that are both technically robust and socially resilient.

### 5.1 Cross-Border Energy Grid Interconnection

Energy networks are among the critical infrastructures, increasingly spanning multiple jurisdictions. From a technological point of view, securing these systems requires, among other things, encrypted SCADA communication, strong cryptographic protocols, Trusted Execution Environments (to protect calculations), and the enforcement of zero trust to prevent unauthorized access. These measures thus create a resilient technical foundation.

However, socio-cultural realities complicate this situation. Operators in different countries often have different training standards, operating cultures, and risk perceptions. Older teams may furthermore resist new security paradigms due to their long-standing dependence on traditional systems, while language barriers and differing regulations hinder collaboration. These factors can undermine even the most advanced technical controls.

### 5.2 Federated Healthcare Data Exchange

An additional example is the healthcare system. These systems are increasingly dependent on secure data exchange for diagnosis and research. From a technological perspective, this includes federated identity systems for cross-institutional authentication, cryptographic guarantees for data confidentiality, and TEEs to ensure data protection in AI-supported analyses. These mechanisms also form the backbone of trustworthiness here.

However, socio-cultural factors also increase complexity in this area. Medical professionals differ in terms of digital literacy, data privacy awareness, and compliance culture. Hospitals in rural areas may have fewer resources than urban facilities, leading to inconsistent security measures. Ethical standards and patient expectations regarding data privacy vary by region and influence the acceptance of new technologies.

### 5.3 Results and Discussion

The analysis of the two use cases illustrates how technological mechanisms and socioeconomic interventions combine to shape trust and trustworthiness in complex, interdependent infrastructures. In both areas, technological controls create verifiable, auditable evidence of trustworthiness, while socioeconomic measures determine whether this evidence is understood, accepted, and integrated into daily practice.

From a technological perspective, both use cases demonstrate that cryptographic security measures, identity systems, zero-trust architectures, and TEEs significantly strengthen the objective security posture of critical systems. In the energy sector, encrypted SCADA communications, robust key management processes, and attested workloads reduce the attack surface and improve data provenance. The enforcement of zero trust replaces implicit trust boundaries with dynamic, request-based authorizations, thereby limiting opportunities for lateral movement. These mechanisms lead to measurable improvements in cryptographic hygiene, interoperability, and forensic reconstructability. Similarly, in healthcare, federated identity systems and context-aware authorizations provide auditable access paths suitable for clinical audits, while TEEs protect sensitive data sets used in AI-powered diagnostic workflows. The result is an interoperable trust structure characterized by non-repudiation, minimized over privileges, and privacy-compliant analytics. In both sectors, the technical layer thus serves as a producer of verifiable evidence -certification proofs, signed telemetry, encryption suite inventories- that underpins institutional trustworthiness.

However, the results also make it clear that technological evidence alone is not sufficient to maintain trust in practice. Socioeconomic factors determine how stakeholders interpret technical controls, whether they are willing to follow new procedures, and how they assess the fairness and legitimacy of system behavior. In the energy use case, intercultural training harmonizes risk perception and reduces misinterpretations of alerts across languages and operating cultures. Standardized frameworks such as IEC 62443 and NIS2 create common expectations across national jurisdictions, while shared governance structures reduce ambiguity about roles and escalation paths. These measures are accompanied by measurable improvements, including fewer cross-border handover errors and a greater willingness to share indicators of compromise. In healthcare, initiatives to promote digital literacy improve the correct use of identity and security tools, reducing misrouting and misuse of emergency measures. Contextual guidelines enable rural hospitals to meet minimum legal requirements despite limited resources, preventing the emergence of unfair, two-tiered trust systems. Transparency materials for patients further improve perceived fairness and increase trust in data sharing and AI-assisted care. These socioeconomic measures thus translate technical trustworthiness into genuine social trust by promoting understanding, incentives, and accountability.

The combined results underscore a key finding: trust in critical digital infrastructures is created by aligning technical safeguards with sociocultural and organizational practices. Technology can determine what is secure, verifiable, and enforceable, but socioeconomic design determines whether these capabilities are accepted, understood, and defended. Sustainable trust therefore arises not solely from technical sophistication, but from the joint development of technical evidence, institutional governance, and human-centered engagement strategies.

**Table 1: Trust & Trustworthy Aspects for Energy Grid and Federated Healthcare Use-Cases**

| Aspect | Cross-Border Energy Grid Interconnection | Federated Healthcare Data Exchange |
|---|---|---|
| **Primary Goal** | Secure and resilient energy distribution across national borders | Secure sharing of sensitive patient data across institutions |
| **Technological Foundations** | SCADA security and IoT integration;<br>Cryptographic protocols;<br>Trusted Execution Environments to control system integrity;<br>Zero Trust for continuous verification: | Federated Identity Systems;<br>Cryptographic guarantees for data confidentiality;<br>Trusted Execution Environments for confidential AI-driven diagnostics; |
| **Socio-Cultural Factors** | Different training standards across countries;<br>Language barriers; | Variability in digital literacy among medical staff;<br>Cultural norms around patient privacy and consent;<br>Resource disparities between rural and urban hospitals; |
| **Trust Gaps and Challenges** | Cascading failure affecting millions;<br>Geopolitical tensions;<br>Dependence on alignment of multiple national authorities;<br>Limited interoperability in legacy SCADA systems;<br>Potential misunderstanding of automated risk scores | Balancing innovation with privacy;<br>Ensuring interoperability without compromising security;<br>Persistent variation in resource availability;<br>Uneven adoption of federated identity tools;<br>Patient skepticism toward algorithmic decision-making; |
| **Trust-Building Mechanisms** | Alignment of cryptographic assurances with standardized cross-cultural operational practices | Alignment of technical privacy protections with clinician competence, hospital resources, and patient expectations |

## 6. Conclusion and Future Work

In an increasingly interconnected and digitized world, trust and trustworthiness have become essential prerequisites for the resilience of critical infrastructures and the security of digital interactions. Yet despite significant advances in cryptographic safeguards, identity management, and zero-trust architectures, a persistent gap remains between technically engineered trustworthiness and the socio-cultural factors that shape how individuals and institutions actually perceive, interpret, and implement trust. Therefore, this work aims to bridge this gap by examining how measurable technical assurances interact with human behavior, cultural norms, and organizational practices across heterogeneous environments.

The analysis shows that technologies such as multi factor authentication, federated identity systems, trusted execution environments, and adaptive zero-trust mechanisms provide strong, verifiable evidence of trustworthiness by reducing attack surfaces, strengthening authentication, and enabling attestation-based assurance. However, these mechanisms only unfold their full effectiveness when complemented by socio-economic measures—including intercultural training, governance alignment, digital literacy programs, and transparent communication strategies—that enable stakeholders to understand, accept, and correctly apply technical controls.

Across all examined use cases, it becomes clear that sustainable trust is not a purely technical outcome but a socio-technical achievement. It emerges only when technical safeguards and human factors are consciously coordinated to reinforce one another, ensuring that engineered trustworthiness is matched by informed and confident user trust.

## Acknowledgment

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 16KIS2239K SUSTAINET_guarDian). The authors alone are responsible for the content of the paper.

**Ethics and Data Use Statement:** This research used publicly available, anonymized secondary data. No identifiable personal information was accessed. Ethical approval was not required in accordance with institutional policies.

**AI Assistance Statement:** The authors used generative AI tools solely for linguistic editing and improving readability (e.g., grammar, phrasing, and structure). No AI system contributed to the formulation of the research questions, methodology, data interpretation, theoretical reasoning, or conclusions. All content was critically reviewed and approved by the authors, who take full responsibility for the manuscript.

## References


Ackerhans, S., Wehkamp, K., Petzina, R., Dumitrescu, D., and Schulz, C., "Perceived Trust and Professional Identity Threat in AI-Based Clinical Decision Support Systems: Scenario-Based Experimental Study on AI Process Design Features", JMIR Publlications – Advancing Digital Health & Open Science, vol. 9, 2025. DOI: 10.2196/64266

Baron, N., Kidd, C., McNealy, J., Rainie, L., "Teens, Trust, and Technology in the Age of AI: Navigating Trust in Online Content", *Common Sense Media*, [online], Available at: https://www.commonsensemedia.org/sites/default/files/research/report/teens-trust-and-technology-in-the-age-of-ai_v2_web.pdf, 2025

Boston Institute of Analytics, "The Biggest Cyber Attacks of 2025: Lessons Learned and the Need for Cybersecurity Experts", [online], Available at: https://bostoninstituteofanalytics.org/blog/the-biggest-cyber-attacks-of-2025-lessons-learned-and-the-need-for-cybersecurity-experts/, [Accessed 2026-01-12]

Bullock, J.B., Pauketat, J.V.T., Huang, H., Wang, Y.-F-, and Anthis, J.R., "Public Opinion and the Rise of Digital Minds: Perceived Risk, Trust, and Regulation Support", *Public Performace & Management Rview, Generative AI and Public Administration: Opportunities and Challenges,* vol. 48, no. 6, 2025, DOI: 10.1080/15309576.2025.2495094

Cho, J.H., Chan, K., and Adali, S., "A Survey on Trust Modeling", *ACM Computing Surveys (CSUR),* vol. 48, no. 2, pp. 1—40, 2015, DOI: 10.1145/281559

Chokkanathan, K., Karpagavalli, S.M., Priyanka, G., Vanitha, K., Anitha, K., and Shenbagavalli, P., "AI-Driven Zero Trust Architecture: Enhancing Cyber-Security Resilience", *8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2024, DOI: 10.1109/CSITSS64042.2024.10816746

Directive (EU) 2022/2555 of the European Parliament and of the Council of 12 December 2022 on measures for a for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), Availbake at: https://eur-lex.europa.eu/eli/dir/2022/2555

Fettweis, G.P., Grünberg, P., Hentschel, T., and Köpsell, S., „Conceptualizing Trustworthiness and Trust in Communications ", *IEEE Communications Magazine*, vol. 63, no. 12, 2025, 10.1109/MCOM.001.2400383.

Finney, G., and Kindervag, J., "The Identity Cornerstone", in Project Zero Trust: A Story about a Strategy for Aligning Security and the Business, Wiley Data and Cybersecurity, pp. 57 – 72, 2023, DOI: 10.1002/9781394255146

Gambo, M.L., and Almulhen, A., "Zero Trust Architecture: A Systematic Literature Review", Journal of Network and System Management, vol. 34., no. 25, pp. 772 – 776, DOI: 10.1109/IAECST57965.2022.10062213

Gerlich, M., "Exploring Motivators for Trust in the Dichotomy of Human—AI Trust Dynamics", *social sciences*, vol. 13, no. 5, 2024, DOI: 10.3390/socsci13050251

Glazer, S., and Karpati, S., „The Role of Culture in Decision Making", *The Cutter Edge IT Journal,* vol. 27, no. 9, pp. 23—29, 2014. Lipps, C., Herbst, J., and Schotten, H.D. "How to Dance your Passwords: A Biometric MFA-scheme for Identification and Authentication of Individuals in IIoT Environments", *16th International Conference on Cyber Warfare and Security (ICCWS),* Cookeville, Tennessee, US, 2021.

Huynh, M.-T., and Aichner, T., „In generative artificial intelligence we trust: unpacking determinants and outcomes for cognitive trust ", *AI & Society,* vol. 40, pp. 5849—5869, 2025, DOI: 10.1007/s00146-025-02378-8

Lipps, C., Weinand, A., Krummacker, D., Fischer, C., and Schotten, H.D., „Proof of Concept for IoT Device Authentication Based on SRAM PUFs Using ATMEGA 2560-MCU", *1st International Conference on Data Intelligence and Security (ICDIS),* South Padre Island, TX, USA, 2018, DOI: 10.1109/ICDIS.2018.00013.

Lipps, C., Mallikarjun, S.B., Strufe, M., Heinz, C., Grimm, C., and Schotten, H.D., „Keep Private Networks Private: Secure Channel-PUFs, and Physical Layer Security by Linear Regression Enhanced Channel Profiles", *3rd International Conference on Data Intelligence and Security (ICDIS)*, Shenzhen, China, 2020, DOI: 10.1109/ICDIS50059.2020.00019.

Lipps, C., Baradie, S., Herbst, J., Armistead, L., and Schotten, H.D., "Cybersecurity in Industrial Automation and Control Systems: The Recent Attack of the Colonial Pipeline", in *Modelling Nation-state Information Warfare and Cyber-operations*, ACPIL, 2022, ISBN: ISBN: 978-1914587382

Luhmann, N., Davis, H. (translated by), Raffan, J. (translated by), Rooney, K. (translated by), King, M. (Editor), Morgner, C. (Editor), "Trust and Power", *Polity*, 2017, ISBN: 978-1509519453

Naprys, E., "Hackers pose as job seekers: opening a resume leads to ransomware" cybernews, [online], Available at: https://cybernews.com/security/hackers-target-recruiters-by-sending-malicious-resumes/, [Accessed 2026-01-15]

Pritchard, S., "The Evolving Cybersecurity Challenge for Critical Infrastructure", *Infosecurity Magazine,* [Online], Available at: https://www.infosecurity-magazine.com/news-features/cybersecurity-for-critical/, [Accessed 2026-01-12]

Pujari, S.R., and Hussain, M.A., "Human Factors in Cybersecurity: Behavioral Insights into Phishing and Social Engineering Attacks", *Nanotechnology Perceptions*, vol. 20, no. S15, ISSN: 1660-6795.

Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act)

Schneier, B., "Applied Cryptography: Protocols, Algorithms, and Source Code in C", Wiley & Sons, ISBN: 978-0471117094, 1995.

Shepherd, C., and Markantonakis, K., "Trusted Excecution Environments", Springer, 2024, ISBN: 978-3031555602

Sithungu, S., and Lipps, C., "Critical Infrastructure Security and the Role of AI: An Overview", 24th European Conference on Cyber Warfare and Security (ECCWS), Kaiserslautern, Germany, 2025, DOI: 0.34190/eccws.24.1.3770TrustTalk, "Trusting, Trustworthiness and Trust", [online], Available at: https://trusttalk.co/trusting-trustworthiness-and-trust/, [Accessed 2026-01-12],

Yates, J.F., and de Oliveira, S., "Culture and decision making", *Organizational Behaivor and Human Decision Process,* pp. 106—118, 2016, 10.1016/j.obhdp.2016.05.003.

Yamamoto, Y., "A morality based on trust: Some reflections on japanese morality", *Philosophy East and West*, vol. 40, no. 4, pp. 451—469, 1990, DOI: 10.2307/1399351

Van Der Ham, J., "Toward a Better Understanding of "Cybersecurity", Digital Threats: Research and Practice, vol. 2, no. 3, pp 1-3, DOI: 10.1145/3442445