Article:
Koghut, M., Lee, S.H. and Al-Tabbaa, O. (Accepted: 2026) Digital Trust: A Multilevel Framework for Trust in Algorithmic Systems. Journal of the Association for Information Systems. ISSN: 1536-9323 (In Press)

Digital Trust:
A Multilevel Framework for Trust in Algorithmic Systems

Maksym Koghut
Business School, Manchester Metropolitan University, Manchester, M15 6BH, UK
m.koghut@mmu.ac.uk

Soo Hee Lee
Kent Business School, University of Kent, Canterbury, CT2 7FS, UK
s.h.lee@kent.ac.uk

Omar Al-Tabbaa
Leeds University Business School, University of Leeds, Leeds, LS2 9JT, UK
busofoa@leeds.ac.uk

Cite as:

Koghut, M., Lee, S.H. and Al-Tabbaa, O. (forthcoming) 'Digital trust: A multilevel framework for trust in algorithmic systems', ***Journal of the Association for Information Systems.***

Digital Trust: A Multilevel Framework for Trust in Algorithmic Systems

Abstract

Organizations now delegate consequential functions to algorithmic systems whose autonomous operations produce outcomes that exceed real-time human reconstruction, even when code and documentation are fully available. Existing trust theory was not built for this. Traditional mechanisms presuppose assessable intentions, human access points, and assignable probabilities; algorithmic environments jointly strain all three. We develop the concept of digital trust, the willingness to accept vulnerability through delegation to algorithmic systems whose operations are not fully graspable and whose behaviors may emerge beyond original design intent. The framework draws primarily on Luhmann's systems theory, with Giddens and Coleman as complements, and turns on a paradox he identified: systems deployed to reduce complexity must themselves generate new complexity. That paradox is sharpened in algorithmic settings through temporal compression, emergent non-deducibility, and coupling at scale. We specify how trust forms across three analytically distinct levels—design, operations, and experience—and identify operative complexity as the residual condition that traditional mechanisms leave unresolved once algorithmic systems begin operating autonomously. Architectural decisions and operational experience shape trust through a dual temporal structure, working through different channels at different speeds. Trust transfers across algorithmic systems through recognition of shared enabling mechanisms rather than surface similarity alone. The argument is illustrated across four algorithmic settings: smart contracts, AI systems, IoT deployments, and platform ecosystems.

Keywords: Digital trust, Algorithmic systems, Complexity paradox, Systems theory, Multilevel framework, Trust mechanisms

Acknowledgements: The authors would like to thank the senior editor, Professor Jan Recker, and the three anonymous reviewers for their constructive comments and suggestions, which greatly strengthened the development of this paper. We also thank Professor Cathy Urquhart for her constructive and insightful feedback, which helped improve the quality of this paper.

1 Introduction

Contemporary organizations increasingly delegate critical functions to algorithmic systems—information systems whose operational logic is outsourced to computational processes that operate with sufficient autonomy to generate consequential outcomes without direct human intervention (Baird & Maruping, 2021). High-frequency trading algorithms execute millions of transactions per second (Arifovic et al., 2022), machine learning models diagnose diseases (Sadr et al., 2025), smart contracts automate inter-organizational agreements without human intermediation (R. R. Chen et al., 2023; Hacker et al., 2023), and platform ecosystems orchestrate economic activity across organizational boundaries (Hanisch et al., 2026; Tiwana et al., 2010).

This pervasive delegation creates a paradox that Luhmann (1979, 1995) identified for social systems generally but that takes on distinctive intensity in algorithmic environments: systems deployed to reduce complexity must themselves develop internal complexity, simultaneously simplifying their users' decision environments while generating new uncertainty through their own operations. Trading algorithms reduce the complexity of market analysis while their interactions produce emergent market dynamics no individual algorithm was designed to create (Demetis & Lee, 2018; Kirilenko et al., 2017). Smart contracts eliminate counterparty discretion risk while generating new complexities around code correctness, oracle dependencies, and governance (Lumineau et al., 2025; Obermeier & Henkel, 2025). Machine learning systems identify patterns invisible to human analysts while creating epistemic inaccessibility where decision processes resist interpretation even with full model access (Burrell, 2016; Rudin, 2019).

Traditional trust mechanisms evolved for conditions where human agents served as primary coordinators (Mayer et al., 1995; Rousseau et al., 1998; Zucker, 1986). Interpersonal trust presupposes assessable intentions and competence; institutional trust depends on human

intermediaries who translate system operations for lay actors (Giddens, 1990); calculative trust presupposes assignable probabilities (Coleman, 1990). Algorithmic systems operating with sufficient autonomy strain these presuppositions jointly: algorithms lack assessable intentions, autonomous execution bypasses human access points, and emergent behavior resists probability assessment (Glikson & Woolley, 2020; Lacity et al., 2025; Vanneste & Puranam, 2024). The field lacks a unified conceptualization capturing how trust operates systemically and across multiple analytical levels when algorithmic systems generate emergent complexity (Jackson & Panteli, 2023; Schuetz et al., 2025).

Our central research question addresses this problem: *How does trust operate in algorithmic systems under conditions of emergent complexity?* We develop a multilevel framework grounded in Luhmann's (1979, 1995) systems theory, complemented by Giddens' (1990) analysis of trust in expert systems and Coleman's (1990) identification of where calculative trust encounters limits. We define *digital trust* as the willingness to accept vulnerability through delegation to algorithmic systems whose operations are not fully graspable and whose behaviors may emerge beyond original design intent, grounded in structured expectations about system behavior and conscious acceptance of associated risks, and manifested in reliant engagement with algorithmically mediated coordination.

This paper makes three contributions to research on trust in algorithmic environments. First, we extend Luhmann's complexity-reduction insight to algorithmic environments, identifying *operative complexity*—uncertainty generated by autonomous algorithmic operations producing emergent outcomes that exceed actors' capacity for real-time comprehension—as a distinctive condition requiring adapted trust mechanisms (Gefen et al., 2003; McKnight et al., 2011; Söllner et al., 2016). Second, we theorize a dual temporal structure of digital trust: design-level architectural decisions and operations-level experience jointly shape trust

formation, extending the complementarity between formal governance and relational trust (McKnight et al., 1998; Poppo & Zenger, 2002) to environments where the operational phase involves autonomous agents subject to emergent drift (Benbya et al., 2020; Ciborra, 2002; Demetis & Lee, 2018). Third, we identify conditions under which trust transfers between algorithmic systems through recognition of shared enabling mechanisms, connecting transfer dynamics to established trust formation processes (Gefen, 2000; McKnight et al., 1998; Schuetz et al., 2025).

2 Theoretical Foundations

Within the broad category of sociotechnical arrangements for producing, processing, and acting on information (Alter, 2008), we focus on *algorithmic systems*: information systems whose operational logic is delegated to computational processes that transform inputs into outputs through defined rules or learned patterns, operating with sufficient autonomy to generate consequential outcomes without direct human intervention at the point of execution (Baird & Maruping, 2021). This encompasses machine learning systems, smart contracts, high-frequency trading algorithms, Internet of Things (IoT) networks, and platform ecosystems. These systems are inherently sociotechnical—designed by human actors, embedded in institutional arrangements, and maintained through ongoing human labor (Orlikowski & Scott, 2008)—yet the trust challenges they create arise from the interaction between algorithmic operations and organizational contexts rather than from technology in isolation. Agency becomes distributed across human and technical actors (Latour, 2005; Murray et al., 2021), with algorithmic systems exhibiting operational autonomy without requiring consciousness or intentionality (Suchman, 2007).

The framework is scoped to *digital* algorithmic systems: while algorithmic decision-making has non-digital precursors in actuarial tables and codified procedures, the conditions that

generate digital trust—operational scale, temporal compression, and emergent drift—arise with digital implementation. Within this class, *agentic IS artifacts* (Baird & Maruping, 2021) denote the subset whose operational-autonomy property we develop in Section 3.1.1; the broader term extends to both adaptive systems of this kind and deterministic systems such as smart contracts.

2.1 Existing Conceptualizations of Digital Trust

Scholars have advanced six distinguishable approaches to digital trust, each illuminating a dimension while leaving fundamental gaps (Table 1).

A first approach extended interpersonal trust frameworks to digital contexts. Wang and Benbasat (2007) and Komiak and Benbasat (2006) applied dimensions of competence, benevolence, and integrity to technological artifacts, treating technology as a medium for trust rather than a distinct trust object. These extensions proved valuable for early digitalization but cannot capture how algorithmic systems transform trust dynamics through operational autonomy and emergent behavior.

A second approach recognized technology as warranting direct trust. McKnight et al. (2011) distinguished trusting people from trusting technology, developing constructs grounded in perceived functionality, helpfulness, and reliability. Lankton et al. (2015) refined this by separating human-like from system-like trust assessments, and Söllner et al. (2016) extended the perspective to a network of trust spanning the IS, its provider, and the broader infrastructure. These contributions recognized technology as a direct trust object yet maintained largely static, individual-level perspectives insufficient for capturing trust across organizational boundaries or over time (Jackson & Panteli, 2023; Schuetz et al., 2025).

A third approach proposed trust substitution through verification. In blockchain contexts, scholars argued that technical properties such as transparency, immutability, and deterministic

execution can replace trust altogether (Beck et al., 2016; Hawlitschek et al., 2018, 2020; Vidan & Lehdonvirta, 2019). De Filippi et al. (2020) distinguished "trust" (reliance on human discretion) from "confidence" (reliance on predictable system behavior), positioning blockchain as a "confidence machine." Subsequent scholarship demonstrates that complete transparency does not eliminate trust: governance challenges, oracle dependencies, and emergent interactions persist beyond what verification can resolve (Hanisch et al., 2026; Lumineau et al., 2025; Obermeier & Henkel, 2025).

A fourth approach treated agency as a variable. Vanneste and Puranam (2024) argue that perceived agency determines whether trust targets the algorithmic system or its human designer. Killoran et al. (2024) propose "potential for harm" as more appropriate than "betrayal" for non-sentient systems. Glikson and Woolley (2020) reveal that trust trajectories differ fundamentally by artificial intelligence (AI) embodiment: robotic AI follows human-like patterns while virtual and embedded AI often exhibits inverted trajectories. These contributions illuminate how agency perceptions shape trust dynamics but remain primarily at the individual level and do not theorize systemic complexity.

A fifth approach identified deductive trust formation. Obermeier and Henkel (2025) introduce "deductive certainty," whereby full knowledge of an algorithmic system's rules enables logical deduction about future behavior. Yet their findings reveal that even when deductive certainty is technically available, most users rely on partial deduction combined with experience-based trust cues—suggesting that transparency does not collapse trust into mere verification.

A sixth approach adopted systems-theoretic perspectives. Lukyanenko et al. (2022) develop a foundational trust framework grounded in Luhmann's systems theory. Lacity et al. (2025), reviewing 214 studies, identify 13 core assumptions about trust across disciplines, noting that

mainstream scholarship presupposes assessable trustees with recognizable agency and that trust-in-technology research has neglected risk as a constitutive condition (see also Schuetz et al., 2025). These contributions move closest to our perspective but have not been extended into a comprehensive framework addressing trust across multiple analytical levels when algorithmic systems generate emergent complexity.

Table 1. Approaches to Digital Trust

Dimension	Trust Extension	Technology as Trust Object	Trust/ Confidence Substitution	Agency-Contingent Trust	Deductive Trust Formation	Systems-Theoretic Trust	Digital Trust (This Paper)
Key contributors	Gefen et al. (2003); Wang & Benbasat (2007); Komiak & Benbasat (2006); Pavlou & Fygenson (2006); McKnight et al. (2002); Koh et al. (2012)	Benbasat & Wang (2005); McKnight et al. (2011); Lankton et al. (2015); Vance et al. (2008); Söllner et al. (2016); Gefen et al. (2008)	De Filippi et al. (2020); Beck et al. (2016); Hawlitschek et al. (2018, 2020); Lumineau et al. (2025); Utz et al. (2023); Vidan & Lehdonvirta (2019)	Vanneste & Puranam (2024); Killoran et al. (2024); Glikson & Woolley (2020); Cabiddu et al. (2022); Korsgaard et al. (2025)	Obermeier & Henkel (2025); Völter et al. (2023)	Lukyanenko et al. (2022); Lacity et al. (2025); Schuetz et al. (2025); Luhmann (1979)	Building on Luhmann (1979, 1995); Giddens (1990); Coleman (1990); Baird & Maruping (2021)
Core assumption	Digital contexts extend interpersonal trust dynamics	Technology possesses assessable attributes warranting direct trust	Technical verification can replace or transform trust into confidence	Perceived agency of the system determines trust dynamics	Rule transparency enables logical deduction about system behavior	Trust serves a systemic function of complexity reduction	Algorithmic systems generate emergent complexity requiring adapted trust mechanisms
Trust object	Human agents mediated by technology	Technology artifact (with perceived functionality, reliability, helpfulness)	Verification mechanism (code, protocol, mathematical proof)	Variable: the system when perceived as agentic; the designer when perceived as tool	The algorithmic rule set and its execution guarantees	Social or technical systems as complexity-reducing mechanisms	Algorithmic system-in-situation (artifact + configurations + data streams + couplings)
Trust formation basis	Interpersonal cues (benevolence, integrity, competence) transferred to digital context	Perceived system attributes (functionality, helpfulness, reliability) assessed through interaction	Technical properties enabling verification (transparency, immutability, deducibility)	Attribution of intentionality and free will to the system	Logical deduction from disclosed rules, complemented by inductive experience	System performance in reducing environmental complexity	Structured expectations formed through design-level architecture and operations-level experience
Role of complexity	Ignored or treated as context	Acknowledged as binary (simple vs. complex systems)	Resolved through technical design	Implicit: agency attributions vary with system complexity	Addressed through rule transparency, though most users rely on partial deduction	Central: trust exists because of and to manage complexity	Central: emergent complexity as the condition creates conditions for digital trust; engagement as the outcome
Temporal perspective	Static or sequential (initial vs. ongoing)	Static attribute assessment	Design-time assurance	Varies with evolving perceptions of system agency	Complementarity between pre-use deduction and in-use induction	Process-oriented but abstractly specified	Dual temporal: design-time architectural trust + ongoing operational experience
Agency treatment	Human agency assumed; technology passive	Technology with human-like or system-like perceived attributes	Agency irrelevant; mechanism eliminates need for trust judgment	Agency as explicit variable determining trust direction	Limited: assumes rule-following without emergent behavior	Systems possess functional (not intentional) agency	Delegated operational autonomy producing emergent outcomes; agency as surprise-capable behavior through drift
Key limitation	Anthropomorphizes technology; insufficient for autonomous systems	Binary agency view; insufficient attention to emergent complexity	Confuses trust transformation with elimination; overlooks residual trust requirements (Obermeier & Henkel, 2025)	Individual-level; does not address systemic complexity or multilevel dynamics	Scope limited to rule-transparent systems; does not address ML opacity or emergent non-deducibility	Abstractly specified; needs contextualization for algorithmic environments	Requires sufficient operational autonomy and emergent complexity as scope conditions

Four recurrent limitations emerge across these approaches. First, most adopt static or sequential perspectives, treating trust as stable rather than co-evolving with algorithmic systems (Schuetz et al., 2025). Second, single-level analysis prevails; individual perceptions receive attention while trust's operation across organizational boundaries remains undertheorized (Jackson & Panteli, 2023; Lacity et al., 2025). Third, complexity receives insufficient analytical attention, overlooking how algorithmic systems paradoxically reduce certain complexities while generating qualitatively new forms (Benbya et al., 2020; Shaikh & Vaast, 2023). Fourth, anthropocentric bias persists in conceptualizing agency, creating difficulties when applied to systems whose consequential operations lack intentions or social identity (Killoran et al., 2024; Lacity et al., 2025; Ramaul et al., 2026). These limitations point toward a theoretical foundation treating trust as a systemic, multilevel, complexity-responsive phenomenon.

2.2 Trust, Complexity, and Uncertainty: Theoretical Foundations

2.2.1 Luhmann's Systems Theory

Niklas Luhmann's (1979, 1995) social systems theory provides a distinctive lens for trust that is well suited to algorithmic environments. Whereas much trust scholarship begins with individual attitudes or relational dynamics, Luhmann begins with a systemic problem: the world presents more possibilities than any actor can process simultaneously, and this excess—what Luhmann terms complexity—demands mechanisms for selective reduction. Trust is one such mechanism. By treating certain future possibilities as if they were settled, trust enables actors to act despite irreducible uncertainty (Luhmann, 1979, 1988).

Several features prove consequential. First, trust is functionally defined: it exists not as a psychological state in isolation but as a mechanism that reduces complexity, enabling engagement with an environment that would otherwise paralyze action. This shifts attention from what trust *is* to what trust *does*. Second, Luhmann (1988) draws a critical distinction

between *trust* (conscious awareness of risk and deliberate engagement despite vulnerability) and *confidence* (unreflective reliance on familiar structures without active risk assessment). This distinction illuminates a range of stances actors adopt toward algorithmic systems, from reflective trust through routine confidence to unreflective reliance characterizing everyday technology use (De Filippi et al., 2020). Third, Luhmann (1995) distinguishes *personal trust* (based on familiarity with individuals) from *system trust* (reliance on abstract functional systems whose operations one cannot directly observe). Algorithmic systems fit neither category: they lack the personal characteristics grounding interpersonal trust, yet they differ from the institutional systems Luhmann analyzed because their operational logic resides in computational processes rather than social norms or communicative procedures.

A paradox central to Luhmann's theory is fundamental to our analysis: systems deployed to reduce complexity must themselves develop internal complexity to do so effectively (Luhmann, 1995). A system too simple to match its environment's variety cannot reduce that environment's complexity for its users. Yet as a system's internal complexity grows, it becomes less transparent to those who depend on it, creating new uncertainty even as it resolves old. This multilevel perspective—in which trust operates simultaneously across different system levels—proves essential as algorithmic systems create trust requirements spanning individual users, organizational processes, and inter-organizational networks.

2.2.2 Complementary Perspectives: Giddens and Coleman

Luhmann was not the only theorist to frame trust as mediating uncertainty. Giddens (1990) and Coleman (1990) locate trust differently while converging on its fundamental function.

Giddens (1990) emphasizes trust in "expert systems" and the disembedding of social relations from local contexts. Modern life requires trusting abstract systems whose internal operations remain beyond lay observation. Crucially, Giddens identifies "access points" where lay actors

encounter expert systems: the bank teller, the physician, the pilot. At these access points, human intermediaries translate system operations, exercise judgment when circumstances demand, and can be held accountable. When algorithmic systems operate with sufficient autonomy, access points may be absent or ineffective: no human intermediary translates a high-frequency trading algorithm's operations in real time; no expert exercises judgment at the moment of smart contract execution. Giddens' framework thus identifies a specific mechanism through which algorithmic autonomy strains traditional trust.

Coleman (1990) formalizes trust as rational calculation: the trustor weighs the probability of gain against the probability of loss, trusting when the expected value is positive. This formalization illuminates conditions under which calculative trust encounters limits. In algorithmic environments, three conditions jointly undermine calculation: emergent behavior reduces the reliability of probability estimates, temporal compression limits the time available for assessment, and coupling effects mean that failure probabilities are not independent across systems. Where Coleman's framework expects a well-defined probability distribution over trustee behavior, algorithmic systems exhibit heavy-tailed distributions where extreme outcomes are more probable than normal distributions suggest (Taleb, 2010).

The differences among these three theorists—in locating trust in systems, expert mediation, or rational calculation—illuminate the distinctive challenges algorithmic environments create. Our framework draws primarily on Luhmann's complexity-management insight, which offers the most general account of why trust is needed, while incorporating Giddens' access-point analysis, which specifies one mechanism through which algorithmic autonomy strains traditional trust, and recognizing the limits Coleman's model encounters when algorithmic operations resist probability assessment.

2.3 Complexity in Algorithmic Environments

Complexity plays a central role in our framework, and its meaning requires specification. Following Luhmann (1995), we understand complexity as a condition where the number of possible states and connections exceeds what can be simultaneously actualized, requiring selective attention and creating contingency. This aligns with Weaver's (1948) concept of "organized complexity": situations involving many interdependent variables whose interactions exhibit partial order rather than simple determinism or pure randomness (see also Miller & Page, 2007). Complexity is relational rather than absolute; it describes the gap between a system's possible states and an observer's capacity to apprehend them. We do not refer to computational complexity in the algorithmic analysis sense, to mere complicatedness (many parts with stable, decomposable relations), or to complexity as a rhetorical intensifier.

Three analytically separable forms of complexity prove consequential for trust in algorithmic environments. The decomposition is analytical rather than ontological: complexity theory treats these dimensions as jointly co-producing complexity through ongoing interaction (Miller & Page, 2007), and the three forms typically co-occur and mutually reinforce one another in algorithmic environments. We separate them to specify how each contributes to the trust challenges algorithmic autonomy generates. *Environmental complexity* refers to volatile, heterogeneous conditions that force organizations to make selections under uncertainty about which possibilities will materialize. *Interaction complexity* arises from coupling across systems, organizational routines, and multiple actors, so that effective behavior cannot be attributed to any single component (Barile et al., 2022; Benbya et al., 2020; Corning, 2002). *Temporal complexity* accumulates through ongoing modifications, versioning, and maintenance, so that a system's current operational state cannot be reduced to its original specification (Zhang et al., 2020).

Emergent complexity—complexity characterized by emergent properties, where macro-level behaviors cannot be deduced from knowledge of component-level rules (Miller & Page, 2007)—characterizes many system types; even simple rule-based systems exhibit such behaviors when components interact. Algorithmic systems, however, generate emergence with three characteristics consequential for trust. First, *temporal compression*: algorithmic operations produce emergent effects at speeds exceeding human real-time assessment, collapsing the temporal window within which traditional trust mechanisms operate (Arifovic et al., 2022). Second, *emergent non-deducibility*: when coupled algorithmic systems produce aggregate behaviors not deducible from individual system specifications, actors cannot form reliable expectations through analysis of individual components (Demetis & Lee, 2018). Third, *coupling at scale*: algorithmic systems routinely interact across organizational, sectoral, and jurisdictional boundaries, meaning that locally manageable complexity aggregates into systemically unmanageable complexity (Benbya et al., 2020).

The distinction is qualitative, not merely scalar. Organizations have long confronted complex environments, yet traditional trust mechanisms rest on epistemic presuppositions that hold in those contexts. Interpersonal trust requires assessable intentions (Mayer et al., 1995); institutional trust depends on access points where human intermediaries translate operations (Giddens, 1990); calculative trust presupposes assignable probabilities (Coleman, 1990). When these presuppositions are jointly undermined by temporal compression, non-deducibility, and coupling at scale, traditional mechanisms encounter structural limits calling for adapted trust mechanisms. Traditional trust remains operative at the *design level*, where human actors make decisions about system adoption, configuration, and governance. Digital trust addresses the *operational* and *experiential* dimensions that provider trust and certification cannot reach, because algorithmic systems generate emergent behaviors in

deployment that no pre-deployment assessment can completely anticipate (Barile et al., 2022; Demetis & Lee, 2018; Zhang et al., 2020).

3 Digital Trust: A Multilevel Framework

3.1 Definition and Core Argument

The theoretical foundations established in Section 2 reveal a theoretical problem: existing trust frameworks, whether grounded in interpersonal assessment, institutional mediation, or calculative reasoning, encounter structural limits when algorithmic systems generate emergent complexity through autonomous operations. We address this problem by defining *digital trust* as: *The willingness to accept vulnerability through delegation to algorithmic systems whose operations are not fully graspable and whose behaviors may emerge beyond original design intent. It is grounded in structured expectations about system behavior and conscious acceptance of associated risks, and manifested in reliant engagement with algorithmically mediated coordination.*

Table 2 locates digital trust against three adjacent constructs using the theoretical dimensions Sections 2.2.1 and 2.2.2 developed. Where Table 1 mapped scholarly approaches to digital trust, Table 2 specifies the construct-level distinctions that motivate our definition: the epistemic basis on which expectations form, the availability of access points through which actors engage trusted systems, the tractability of calculative assessment, and the conditions under which each trust form encounters structural limits. The rightmost column locates, through illustrative cases, where the boundary between digital trust and its adjacent constructs falls along each dimension.

Table 2. Digital Trust and Adjacent Trust Constructs

Dimension	Interpersonal Trust	System Trust	Confidence	Digital Trust (this Paper)	Boundary and Illustrative Cases
Functional problem addressed	Social complexity arising from human unpredictability (Mayer et al., 1995; Luhmann, 1979)	Institutional complexity arising from dis-embedded social relations (Giddens, 1990; Luhmann, 1979)	Routine complexity absorbed through familiarity with stable environments (Luhmann, 1988)	Operative complexity arising from autonomous algorithmic operations producing emergent outcomes beyond actors' real-time comprehension	Boundary: the source of complexity. A patient's reliance on a physician to interpret medical knowledge rests on system trust (institutional complexity mediated by expert access points); reliance on an autonomous AI diagnostic system operating without clinician mediation requires digital trust (operative complexity generated within the system itself).
Epistemic basis for expectations	Theory of mind; assessment of intentions, competence, and character (Mayer et al., 1995)	Institutional guarantees; expert credentials; regulatory oversight (Giddens, 1990)	Unreflective familiarity with routine patterns; expectations confirmed by repetition (Luhmann, 1988)	Design-level architectural parameters combined with operations-level experiential calibration under epistemic inaccessibility	Boundary: what grounds expectation formation. A frequent traveler's expectations about an airline draw on institutional guarantees and regulatory oversight (system trust); a DeFi user's expectations about a smart contract's execution calibrate design-level code audits against operations-level observation under epistemic inaccessibility (digital trust).
Access point structure (Giddens, 1990)	Direct interaction with the trusted agent	Human intermediaries translate system operations at defined encounter points	No access points needed; reliance is pre-reflective	Access points absent or ineffective; algorithmic operations bypass human intermediaries at execution	Boundary: whether a human intermediary is positioned at execution. Commuters' reliance on a conventionally operated rail line rests on system trust - the conductor and dispatcher are access points; reliance on a fully autonomous transit system where algorithmic control executes without on-site human intervention requires digital trust.
Calculative tractability (Coleman, 1990)	Probability estimates feasible from observable behavior, reputation, and relational history	Probability estimates supported by institutional track records and normative frameworks	Calculation unnecessary; outcomes taken for granted	Probability estimates undermined by temporal compression, emergent non-deducibility, and coupling at scale	Boundary: whether conditions permit stable probability estimates. An underwriter's actuarial assessment of default risk fits calculative trust; a risk manager's engagement with a machine learning model's failure modes - where tail events emerge from coupling and drift - exceeds calculation and requires digital trust.
Risk awareness (Luhmann, 1988)	Conscious; vulnerability to betrayal, incompetence, or defection	Conscious; vulnerability to institutional failure or expert error	Prereflective; risk not actively assessed until disruption triggers reflection	Conscious; vulnerability to emergent behaviors, systemic cascades, and consequences of epistemic inaccessibility	Boundary: whether risk is tacitly absorbed through routine or consciously engaged under emergent uncertainty. An employee's daily use of standard office software rests on confidence (risk unreflective until disruption); a CFO's authorization of an algorithmic trading strategy requires digital trust, with emergent cascade risks consciously accepted.
Scope condition	Holds where intentions and competence are assessable through interaction or reputation	Holds where functioning access points connect lay actors to expert systems	Holds where routines remain undisrupted and expectations are continuously confirmed	Applies where operational autonomy generates emergent complexity that jointly undermines the presuppositions of the three adjacent forms	Boundary: where operational autonomy generates emergent complexity that jointly undermines the presuppositions of interpersonal, institutional, and calculative trust. A conventional ERP system under continuous human oversight falls within traditional trust scope; a supply chain orchestrated by autonomous agents whose coupling produces emergent behavior falls within digital trust scope.

The definition preserves vulnerability as constitutive of trust (Mayer et al., 1995; Rousseau et al., 1998) while specifying the conditions—epistemic inaccessibility and emergent drift—that distinguish digital trust from adjacent forms. As Table 2 shows, each adjacent construct rests on an epistemic presupposition that algorithmic environments strain: interpersonal trust requires assessable intentions, system trust depends on functioning access points, and calculative trust presupposes assignable probabilities. When operational autonomy jointly undermines all three, digital trust addresses the residual uncertainty. The boundary delineating digital trust is functional rather than ontological. The framework applies wherever operational autonomy generates emergent complexity that bypasses the access points on which traditional trust depends (Giddens, 1990), regardless of substrate. A rail system governed by autonomous algorithmic control falls within scope; a conventionally operated rail system does not. The differentiating criterion is not the technology itself but whether autonomous operations produce emergent complexity jointly undermining the epistemic presuppositions of interpersonal, institutional, and calculative trust.

Figure 1 represents the framework's core logic: *Algorithmic System Characteristics* → *Emergent Complexity* → *Digital Trust* → *Productive Engagement*. Algorithmic systems reduce operational complexity through standardization, automation, and information processing at scale, enabling organizations to manage previously impossible scales of activity. Yet these same characteristics generate new complexities: interaction effects create exponential coupling possibilities, emergent behaviors arise from system interactions, and governance challenges multiply as algorithmic operations outpace oversight.

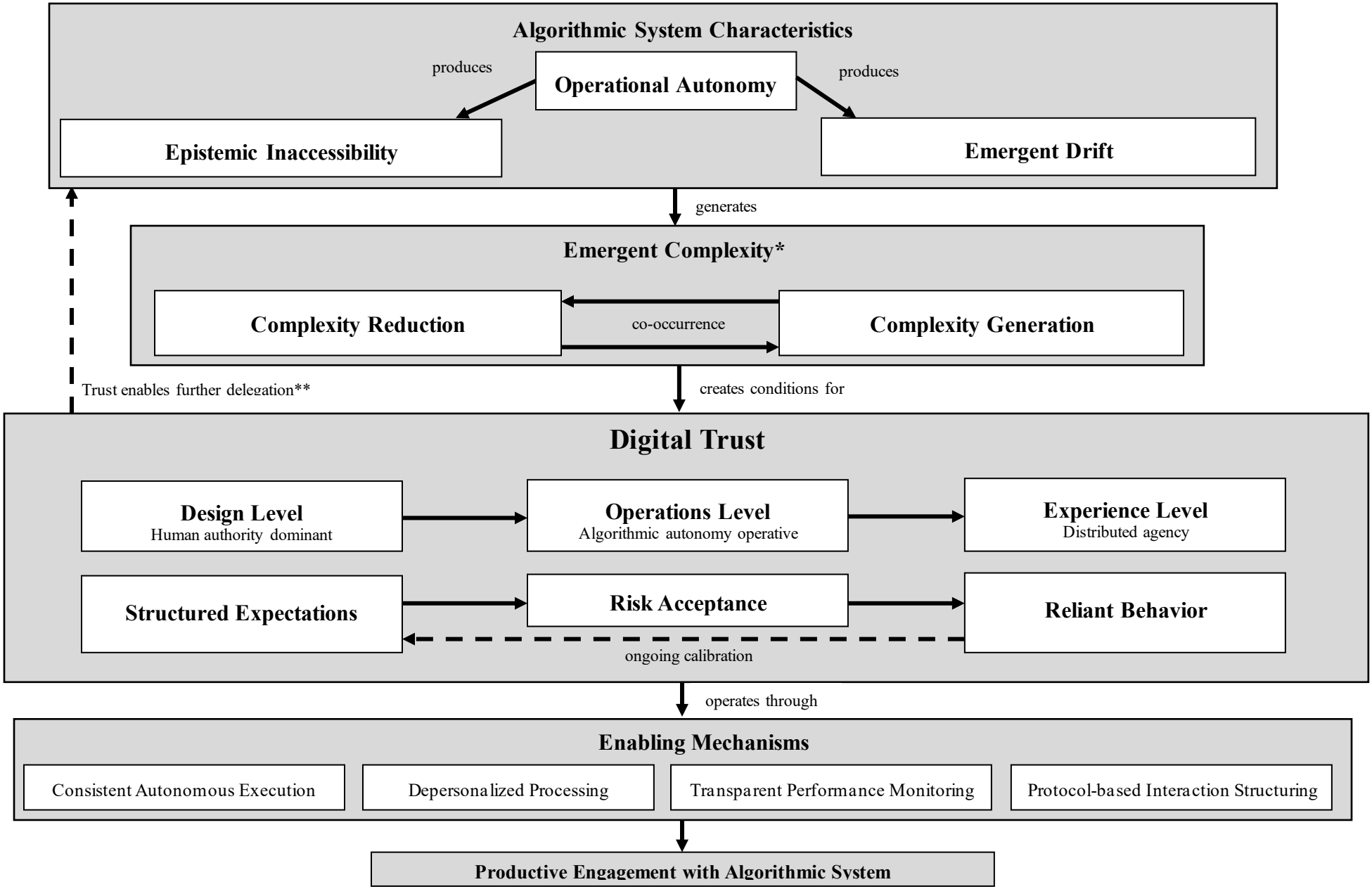

Figure 1. A Conceptual Framework for Digital Trust in Algorithmic Systems

*Notes: *Luhmann's paradox: systems deployed to reduce complexity simultaneously simplify and generate new uncertainty. **The dashed arrow indicates that established digital trust enables further delegation to algorithmic systems, which may generate additional complexity requiring ongoing trust calibration.*

Digital trust emerges as the mechanism enabling organizations to function despite this paradoxical complexity. It integrates architectural trust (established through design-level decisions about system parameters) with processual trust (evolving through operational experience), creating a form of engagement that neither simple extension of interpersonal trust nor mere confidence in technical reliability can capture. One feedback relationship shapes ongoing trust dynamics: established digital trust enables deployment of more sophisticated algorithmic systems, as organizations confident in their capacity to navigate algorithmic complexity invest in more autonomous, interconnected systems. This can generate new complexity requiring further trust adaptation. Trust failures, conversely, can trigger resistance to algorithmic advancement.

Organizations may respond to algorithmic complexity through means other than digital trust, including avoidance, resistance, or unreflective reliance. The framework identifies digital trust as a theoretically coherent response, not an inevitable outcome. Equally, prior trust—in a technology provider, institutional setting, or regulatory regime—may inform initial adoption decisions; the framework does not preclude this but focuses on what happens once algorithmic systems begin operating with sufficient autonomy to generate emergent complexity.

3.1.1 Algorithmic System Characteristics

The framework characterizes algorithmic systems through one primary characteristic and two consequential properties (Figure 1, top tier).

Operational autonomy—the capacity to enact consequential decisions without direct human intervention at execution (Baird & Maruping, 2021)—is the primary characteristic distinguishing systems that require digital trust from those manageable through traditional mechanisms. Systems executing precisely specified instructions under continuous human supervision do not generate the emergent complexity the framework addresses. As autonomy increases, systems gain the capacity to produce outcomes their designers did not anticipate, creating conditions where traditional trust mechanisms encounter structural limits.

Operational autonomy at sufficient scale and in coupled environments generates two consequential properties. First, *epistemic inaccessibility*: autonomous systems operating at scale and speed produce decision processes that exceed human cognitive capacity for real-time reconstruction (Burrell, 2016; Humphreys, 2009). This occurs not because processes are hidden but because they surpass human apprehension. Even fully documented, transparent code becomes epistemically inaccessible when it processes millions of interactions per second. Obermeier and Henkel (2025) demonstrate empirically that even when full code transparency technically enables deductive certainty about a smart contract's execution, only 5% of users conduct thorough security audits—the technical possibility of complete verification does not translate into comprehensive deductive analysis at scale.

Second, *emergent drift*: autonomous systems in complex environments produce behaviors diverging from design intent through coupling with changing inputs, other systems, and user adaptations (Ciborra, 2002; Pentland et al., 2020). Even perfect initial understanding of a system's logic does not guarantee understanding of its behavior over time, because performance is produced in deployment through interactions designers did not fully anticipate. Drift arises from exogenous changes in the data environment, endogenous modifications through updates and patching, and interaction effects among coupled systems

creating untested execution paths. Adaptive computational processes are especially susceptible, whereas deterministic ones suppress drift by construction (Pentland et al., 2020). The system becomes a moving target; inspecting it once does not stabilize understanding.

3.2 The Complexity Paradox in Algorithmic Environments

3.2.1 Complexity Reduction

Algorithmic systems achieve complexity reduction through three mechanisms. *Standardization* eliminates variation by establishing uniform data formats, communication protocols, and processing rules, enabling automated coordination across thousands of organizational partners (Hanseth & Lyytinen, 2010). *Automation* removes human variability from routine operations: trading algorithms execute based on programmed logic, eliminating interpretive differences characterizing human traders (MacKenzie, 2019). *Information processing* at scales exceeding human cognitive limits enables pattern identification across millions of variables invisible to human analysis (Fourcade & Healy, 2017). Together, these mechanisms allow organizations to manage activity at scales that would otherwise overwhelm human coordinative capacity.

3.2.2 Complexity Generation

The same operational autonomy that enables complexity reduction simultaneously generates emergent complexity through three dynamics. *Interaction effects* arise as algorithmic systems couple with other systems, organizational routines, and human actors both within and across organizational boundaries, producing aggregate behaviors no single component designed or controls (Benbya et al., 2020). The cross-boundary case proves especially consequential: when supply chain algorithms from competing firms interact, or when decentralized finance (DeFi) protocols compose across platforms, locally manageable complexity aggregates into systemically unmanageable forms attributable to no individual system. *Emergent behaviors* exceed design specifications as algorithmic systems interact with dynamic environments: the

2010 Flash Crash exemplifies this, as interacting trading algorithms produced a market collapse no individual algorithm was designed to cause (Demetis & Lee, 2018; Kirilenko et al., 2017). *Cross-boundary coordination* intensifies as algorithmic operations outpace human oversight capacity, requiring novel accountability frameworks for outcomes emerging from system interactions rather than individual decisions (L. Chen et al., 2022; Lumineau et al., 2021).

3.2.3 The Paradox as Irreducible

This paradox cannot be resolved through technological advancement. Machine learning systems that handle complex patterns create new complexities around interpretability and model drift (Cremer et al., 2019). Blockchain systems that reduce transaction trust requirements create new complexities around consensus and governance (Li et al., 2017; Lumineau et al., 2021, 2025). More sophisticated systems resolve certain complexities while inevitably producing others.

The irreducibility clarifies why digital trust is functional. Organizations cannot eliminate algorithmic complexity; they must develop mechanisms for acting productively despite it. Digital trust serves precisely this function: it enables engagement with systems whose full operational behavior cannot be comprehended, by forming workable expectations, accepting residual risks, and sustaining reliant engagement.

3.3 Three Levels of Analysis

The framework distinguishes three interconnected but analytically distinct levels—design, operations, and experience—reflecting that algorithmic systems simultaneously exist as human-designed artifacts, operationally autonomous systems, and objects of trust relationships. The design level encompasses two analytically distinct trust dynamics: traditional trust informing adoption and provider selection, and architectural trust establishing

the parameters—code logic, governance protocols, oversight mechanisms—within which operations-level and experience-level digital trust forms.

3.3.1 Design Level

The design level encompasses foundational structures: software architectures, governance mechanisms, regulatory frameworks, and the human organizations that create, deploy, and maintain algorithmic systems. Human agency dominates. Humans write code, select architectures, establish protocols, and retain ultimate authority to modify or discontinue systems (Rahwan, 2018). Legal frameworks operate primarily here; the EU's AI Act mandates specific requirements including human oversight mechanisms and technical documentation (Veale & Zuiderveen Borgesius, 2021). In blockchain systems, the design level manifests through consensus mechanism selection, validator configuration, and governance processes for protocol updates. In AI systems, it encompasses model architecture choices, training data curation, and oversight mechanisms.

The Decentralized Autonomous Organization (DAO) incident of 2016 illustrates the interplay between design and operations. The vulnerability exploited was a design-level issue: code permitting reentrancy that auditors failed to anticipate. The exploit was an operations-level event: the system executed exactly as coded. Ethereum's hard fork was a design-level intervention demonstrating that human authority over design can override operational outcomes when communities mobilize (DuPont, 2017). The controversy arose because the intervention violated expectations of operational autonomy, revealing the tension between design-level control and operations-level autonomy at the heart of digital trust.

3.3.2 Operations Level

The operations level represents where algorithmic systems function according to programmed logic and learned patterns. Once deployed, systems process inputs, execute operations, and produce outputs with varying degrees of autonomy. At this level, algorithmic systems exhibit

emergent properties not predictable from design specifications alone. Machine learning systems develop internal representations defying human interpretation (Pasquale, 2015). Rule-based systems coupled across organizational boundaries produce interaction effects no single specification anticipated. A recommendation algorithm processes millions of user interactions, adapting through feedback loops no human directly controls. Trading systems execute thousands of transactions per second, responding to conditions faster than human cognition permits (Arifovic et al., 2022). The operations level generates the specific trust challenges distinguishing digital trust from trust formed purely at the design level through provider reputation or certification.

The operations level is where epistemic inaccessibility and emergent drift become concrete rather than theoretical. The behavioral patterns that emerge provide the empirical foundation for trust calibration: actors develop expectations from observed algorithmic performance rather than from design specifications, creating a gap between design intent and operational reality that the experience level must bridge.

3.3.3 Experience Level

The experience level represents where humans develop, maintain, and modify trust through interaction with operational algorithmic systems. Trust forms not through direct access to design details or complete operational understanding but through accumulated experiences, observed patterns, and social influences (Cabiddu et al., 2022; Glikson & Woolley, 2020). People apply different evaluative standards to algorithmic versus human decisions, often exhibiting "algorithm aversion" after observing errors even when algorithms outperform humans on average (Burton et al., 2020; Dietvorst et al., 2015, 2018). Conversely, "algorithm appreciation" emerges when users prefer algorithmic judgment for tasks perceived as objective (Logg et al., 2019). Trust can persist despite operational failures if users attribute problems to correctable design-level errors; trust can erode in well-functioning systems if

users develop concerns about privacy, fairness, or control. Social influences—media coverage, peer experiences, cultural attitudes—shape trust formation in ways diverging from operational realities (Cave & Dihal, 2020).

3.3.4 Levels, Agency, and Temporal Dynamics

This multilevel framework resolves a persistent difficulty in theorizing algorithmic agency. Rather than debating whether algorithms "have" agency, the framework locates different forms of consequential action at different levels (see also Korsgaard et al., 2025, for a convergent argument across the trust literature more broadly). At the design level, human agency dominates. At the operations level, algorithmic systems exhibit operational autonomy, producing consequential outcomes according to their own logic within designed parameters. At the experience level, agency becomes distributed as humans and systems jointly shape outcomes—when a credit scoring algorithm denies a loan, it exercises operational autonomy within human-designed parameters; when applicants modify behavior to improve scores, they demonstrate intertwined human and algorithmic action (Kiviat, 2019).

The levels operate on different timescales. Design changes occur slowly, requiring significant coordination. Operational patterns shift more rapidly as systems interact with dynamic environments and drift accumulates. Experience-level trust builds gradually through positive interactions but can erode rapidly through negative events or external revelations (Dietvorst et al., 2015; Schmidt et al., 2020). These different timescales demonstrate how human control and algorithmic autonomy coexist: design-level authority is not negated by operational-level autonomy, and experience-level trust is determined by neither alone.

3.4 Components of Digital Trust

Digital trust comprises structured expectations, conscious risk acceptance, and reliant engagement, following the logical ordering from beliefs through evaluation to behavior that Fishbein and Ajzen (1975) articulated and McKnight et al. (1998) applied to trust formation.

3.4.1 Structured Expectations

Expectations represent beliefs actors develop about algorithmic system behavior. Unlike expectations about human agents, which draw on theory of mind and character assessment (Mayer et al., 1995), algorithmic expectations require reasoning about systems whose internal operations may be epistemically inaccessible (Shin, 2021). Expectations form along four dimensions. *Systemic predictability* encompasses beliefs about consistency and reliability—approaching certainty for deterministic systems like smart contracts, becoming statistical for adaptive systems like machine learning (Jacovi et al., 2021). Even when technical transparency enables deductive certainty, most users form expectations through inductive generalization from observed performance (Obermeier & Henkel, 2025). *Functional integrity* captures expectations that a system will maintain operational coherence without corruption, degradation, or unauthorized modification. *Complexity absorption capacity* addresses beliefs about a system's operational domain—what it can and cannot handle. *Communicated performance* concerns expectations shaped by system outputs including dashboards, alerts, and explanations.

3.4.2 Risk Acceptance

Risk acceptance denotes the conscious decision to engage with algorithmic systems despite recognized uncertainties. This extends Luhmann's (1988) emphasis on risk awareness as constitutive of trust. Algorithmic environments create a distinctive risk profile spanning technical risks (vulnerabilities, failure cascades, adversarial attacks), economic risks (market manipulation, systemic cascades enabled by algorithmic speed), and social risks (bias amplification, autonomy erosion). Actors must accept not only known risks but unknown unknowns—emergent behaviors no analysis could anticipate (Selbst et al., 2019).

3.4.3 Reliant Engagement

Reliant engagement is the behavioral manifestation of digital trust: observable actions demonstrating acceptance of vulnerability to algorithmic operations. In algorithmic contexts,

reliance often involves irreversible commitment—funds locked in protocols, strategies dependent on algorithmic execution, operations premised on system availability (Y. Chen & Bellavitis, 2020). Reliance depth varies from superficial engagement to existential dependence. Organizations restructuring core processes around AI demonstrate deep reliance; casual recommendation engine users demonstrate superficial reliance. Behavioral adaptation reveals how reliance shapes action: actors modify behaviors to align with algorithmic logic, from SEO optimization to credit score management (Zerilli et al., 2019).

3.4.4 Integration and Calibration

These components constitute digital trust through dynamic interaction. Expectations shape risk perception: high predictability expectations make minor deviations appear as major threats. Risk acceptance enables reliant engagement, though actual dependence often exceeds initial acceptance as organizations progressively embed algorithmic dependencies (Cui et al., 2026). Reliant engagement generates observations that confirm or challenge expectations, creating ongoing calibration. Complexity reduction is thus the functional motivation for trust—the reason actors form expectations, accept risks, and sustain engagement—while reliant engagement is its behavioral manifestation, not an independent antecedent of complexity reduction. When expectations align with operational experience and accepted risks match actual vulnerabilities, trust stabilizes. Misalignment drives adaptation: unexpected behaviors force expectation revision, unanticipated risks demand expanded acceptance thresholds, excessive reliance may prompt organizational adjustment (Toreini et al., 2020).

3.5 Enabling Mechanisms

Four mechanisms enable digital trust formation. These mechanisms provide epistemic and operational footholds for expectation formation, risk assessment, and sustained

engagement—they do not replace trust with certainty but supply partial grounds for trust under irreducible uncertainty (Obermeier & Henkel, 2025).

Consistent Autonomous Execution serves as the foundational mechanism, creating predictability through reliable operation independent of human intervention at execution. When smart contracts execute based on programmed conditions, or trading algorithms respond to market signals according to learned patterns, they produce behavioral regularity supporting expectation formation (Baird & Maruping, 2021; Gregory et al., 2024). This regularity does not constitute certainty: emergent drift means past consistency does not guarantee future behavior, particularly as systems interact with dynamic environments. Trust forms because behavioral regularities are sufficient for actors to form workable expectations and accept residual risk.

Depersonalized Processing removes particularistic human characteristics from operational decisions. Algorithmic systems apply consistent rules regardless of participant identity or relational history (Beck et al., 2018), enabling trust precisely by removing the interpersonal dynamics traditional models presuppose. Actors trust a procedure's consistency rather than a person's judgment (Koh et al., 2012). Depersonalization simultaneously removes contextual adaptability, creating tension between procedural fairness and substantive justice when systems cannot accommodate unanticipated circumstances (Selbst et al., 2019).

Transparent Performance Monitoring provides mechanisms for observing and verifying algorithmic operations, creating empirical foundations for expectation calibration. Blockchain's permanent transaction records exemplify radical transparency; machine learning systems increasingly incorporate explainability features (Shin, 2021). Yet transparency is not a monotonic good: raw algorithmic data often overwhelms rather than enlightens, and effective transparency requires translation mechanisms converting technical operations into

actionable understanding (Ananny & Crawford, 2018; Jacovi et al., 2021; Langer et al., 2021). Transparency enables trust formation not by making systems fully comprehensible but by providing selective access points through which actors can partially verify expectations—echoing Giddens' (1990) insight while demonstrating the persistent gap between access and grasp.

Protocol-Based Interaction Structuring creates standardized patterns for engaging with algorithmic systems. API specifications, consensus mechanisms, governance frameworks, and platform rules encode coordination norms directly, substituting for traditional intermediaries (de Reuver et al., 2018; Tiwana, 2013). Protocols enable trust by constraining the space of possible system behaviors, creating bounded uncertainty more amenable to trust formation than open-ended contingency. Protocol governance—versioning policies, deprecation notices, backward compatibility commitments—shapes the temporal dimension of trust by providing assurance about system evolution (Hanseth & Lyytinen, 2010).

These mechanisms operate in combination. A blockchain-based smart contract combines consistent autonomous execution (deterministic code) with depersonalized processing (identity-agnostic rules), transparent monitoring (public ledger), and protocol-based structuring (consensus governance). An AI diagnostic system combines consistent execution (reliable pattern recognition) with transparent monitoring (explainability features) but offers less depersonalization (outputs vary by patient characteristics) and different protocol structuring (clinical governance). The specific configuration shapes the form digital trust takes in each context.

3.6 Boundary Conditions

Three conditions determine when the framework provides analytical value (Whetten, 1989). First, the framework presumes algorithmic systems operating with sufficient autonomy to

generate emergent complexity; simple rule-based systems under continuous supervision may not require the apparatus we develop (Rai et al., 2019). Second, the framework assumes contexts where trust represents a meaningful response to irreducible uncertainty—where stakes are significant and trust enables action otherwise paralyzed by complexity (Luhmann, 1979; Shapiro, 1987). Third, the framework presumes environments where actors maintain meaningful choice about algorithmic engagement; in contexts of monopoly or regulatory mandate, frameworks emphasizing power, compliance, or resistance prove more illuminating (Kellogg et al., 2020; Möllering, 2006). These conditions represent gradients rather than binary boundaries.

4 Illustrating the Theory

Four contexts illustrate the framework's reach across diverse algorithmic configurations: smart contracts, artificial intelligence (AI) systems, Internet of Things (IoT) networks, and platform ecosystems. Each generates the complexity paradox in distinctive form while sharing the framework's common structure—emergent complexity addressed through structured expectations, risk acceptance, and reliant engagement operating across design, operations, and experience levels. Table 3 summarizes the comparison; strength ratings on the enabling mechanisms are relative across the four contexts rather than absolute. Sections 4.1 through 4.4 develop each context in turn, and Section 4.5 draws the synthesis.

4.1 Smart Contracts

Smart contracts provide a revealing context because they are deployed in environments where traditional trust mechanisms are absent or insufficient (Hacker et al., 2023; Koghut et al., 2021). As self-executing programs on blockchain platforms, they automatically enforce agreement terms when predetermined conditions are met. Halaburda et al. (2024) distinguish "strong" smart contracts (prohibitively expensive to adjust ex post) from "weak" smart contracts (adjustable, with automatic execution). With strong smart contracts, trust concerns

center on code correctness and immutability; with weak smart contracts, concerns extend to opportunistic behavior through contract adjustment, since code does not fully specify all contingencies. Validators execute contracts as defined in both cases; the trust difference lies in what the contract governs versus what remains subject to human discretion.

Table 3. Digital Trust Across Four Algorithmic Contexts

Framework Element	Smart Contracts	AI Systems	IoT Systems	Platform Ecosystems
Primary complexity pattern	Deterministic execution eliminates counterparty and settlement risk; generates complexities around code correctness, oracle dependencies, and governance (Lumineau et al., 2025). Strong smart contracts amplify immutability risk; weak smart contracts introduce discretion risk (Halaburda et al., 2024).	Pattern recognition exceeds human analytical capacity; generates epistemic inaccessibility and emergent drift through bias and distribution shift (D'Amour et al., 2022; Rudin, 2019).	Devices automate physical interactions; aggregation produces emergent complexity around interoperability, security, and cross-device coordination (Gubbi et al., 2013). Drift arises through coupling, firmware updates, and environmental change even without adaptive learning.	Common interfaces abstract financial, computational, and informational complexity; generate versioning challenges, dependency risks, and systemic propagation across interconnected services (Constantinides et al., 2018).
Expectations (dominant dimension)	*Deterministic predictability*. Strong smart contracts approach certainty through code-enforced execution; weak smart contracts admit transparent modification. Both involve expectations about functional integrity and communicated performance (Obermeier & Henkel, 2025).	*Probabilistic reliability*. Beliefs about statistical accuracy across populations rather than certainty in individual cases (Yin et al., 2019); most users form expectations through inductive generalization even where transparency is available (Obermeier & Henkel, 2025).	*Physical-digital reliability*. Sensor accuracy, actuator responsiveness, and resilience to environmental interference alongside algorithmic behavior (Litman, 2025). Physical consequence visibility shapes calibration.	*Contractual stability*. API stability, backward compatibility, and consistent performance across versions (Trienekens et al., 2004).
Risk acceptance (dominant concern)	*Code vulnerability and irreversibility*. Strong smart contracts: bugs unpatchable post-deployment. Weak: counterparty opportunism where code underspecifies contingencies (Halaburda et al., 2024). DeFi users assess audits and protocol metrics (Bourveau et al., 2024); enterprise participants rely on institutional relationships, , with governance misalignments - not technical failure - as a primary driver of ecosystem breakdown (Hanisch et al., 2026).	*Epistemic opacity and bias*. Decision processes resist interpretation; outcomes may be unexplainable or biased (Kleinberg et al., 2018). Emergent risks exceed prior analysis (Selbst et al., 2019).	*Physical safety and cascading failure*. Irreversibility of physical action amplifies stakes: smart city traffic systems and medical devices extend risk beyond digital uncertainty to bodily harm (Singh et al., 2020).	*Platform dependency and lock-in*. Service disruption, unilateral terms changes, and costly adaptation to platform evolution (Wareham et al., 2014); progressive integration escalates exposure.
Reliant engagement (dominant form)	*Irreversible commitment*. Billions locked in DeFi protocols (Lin et al., 2024); enterprises embed private blockchains in supply chain coordination. Engagement depth varies sharply between DeFi and enterprise contexts.	*Graduated collaboration*. Augmentation rather than replacement: physicians use diagnostic AI as second opinion; loan officers review algorithmic recommendations (Lebovitz et al., 2022). Reliance deepens as calibration accumulates (Cui et al., 2026; Lebovitz et al., 2022).	*Embodied reliance*. Daily physical encounters with smart home systems; observed operational improvements in industrial IoT (Boyes et al., 2018). Physical consequences render vulnerability tangible.	*Architectural embedding*. API experiments evolve into comprehensive platform strategies; switching costs make reliance increasingly consequential over time (Jacobides et al., 2018).
Design-level features	Consensus mechanism, validator architecture, and governance processes for protocol updates set trust parameters (Renwick & Gleasure, 2020). The DAO incident illustrates design vulnerability → operations event → contested design intervention (DuPont, 2017).	Training data, model architecture, deployment parameters, and oversight shape but do not determine operational dynamics (Burrell, 2016). Regulatory frameworks (EU AI Act) address design but cannot govern operations-level	Device standards, firmware governance, interoperability protocols, and security architectures set parameters. Multi-manufacturer coordination shapes the scope of operational interactions.	API specifications, platform rules, boundary resources, and ecosystem governance establish the interaction space (Tiwana, 2013). Versioning policies and backward compatibility commitments structure temporal dynamics (Hanseth &

Framework Element	Smart Contracts	AI Systems	IoT Systems	Platform Ecosystems
		emergence.		Lyytinen, 2010).
Operations-level dynamics	Deterministic execution within coded parameters; high autonomy across millions of executions (Halaburda et al., 2024). Emergence arises from cross-contract interactions, oracle dependencies, and composability (Goldsby & Hanisch, 2022; Hanisch et al., 2026).	Adaptive learning produces behaviors not explicitly programmed. Black-box dynamics and distribution shift drive drift through exogenous, endogenous, and interaction pathways (D'Amour et al., 2022); adaptive programming is especially prone to such dynamics, whereas deterministic programming is not (Pentland et al., 2020).	Distributed autonomy within coupled networks. Drift arises from coupling, updates, and environmental change even for rule-based devices. Individually comprehensible devices aggregate into system-level non-deducibility.	Lower operational autonomy than AI or smart contracts; interaction effects among heterogeneous participants generate emergent dynamics. Modular architecture propagates change unpredictably across services (Constantinides et al., 2018).
Experience-level patterns	Bifurcated. DeFi users assess audits, protocol metrics, community signals (Bourveau et al., 2024); enterprise participants rely on institutional relationships. Dynamics differ sharply between public and private deployments.	Trust often starts high and declines with experience, inverting interpersonal trajectories (Glikson & Woolley, 2020). Algorithm aversion persists even when AI outperforms humans; minimal modification capacity mediates trust (Dietvorst et al., 2018).	Trust builds through embodied daily encounters rather than abstract interaction (Boyes et al., 2018). Tangible failure (e.g., autonomous vehicle malfunction) accelerates erosion relative to digital-only contexts (Krueger et al., 2025).	Progressive integration deepens trust while raising exit costs. Calibration occurs through cumulative experience with API reliability; ecosystem-level events (e.g., sudden API deprecation) cascade across dependent organizations.
Enabling mechanism configuration				
Consistent autonomous execution	**Strongest**. Deterministic code delivers the highest behavioral regularity (Baird & Maruping, 2021).	**Strong**. Reliable pattern recognition provides statistical consistency; probabilistic rather than deterministic.	**Strong at device level; variable at system level**. Individual devices execute reliably; system-level consistency depends on coupling dynamics.	**Strong**. Standardized interfaces deliver predictable responses; consistency depends on provider governance.
Depersonalized processing	**Strongest**. Identity-agnostic rules treat all inputs by consistent criteria (Beck et al., 2018).	**Moderate**. Outputs vary by input characteristics (patient data, credit profiles); depersonalization is procedural rather than identity-agnostic (Selbst et al., 2019).	**Moderate to strong**. Rule-based behavior applies uniformly; physical context introduces outcome variability.	**Strong**. Rules apply uniformly across participants; tiered access introduces particularism.
Transparent performance monitoring	**Strongest**. Permanent public ledger makes every operation auditable (Shin, 2021).	**Weakest**. Black-box opacity resists meaningful monitoring; explainability features remain limited. Access-grasp gap widest (Langer et al., 2021).	**Moderate**. Sensor logs and telemetry provide observability; distributed architecture limits system-level monitoring.	**Moderate**. Dashboards and SLAs provide selective transparency; platform internals (ranking, matching) remain proprietary.
Protocol-based interaction structuring	**Strong**. Consensus mechanisms and governance frameworks encode coordination norms directly (de Reuver et al., 2018).	**Variable**. Clinical governance and regulatory frameworks provide external structuring; interaction protocols less standardized.	**Critical but fragmented**. Interoperability standards essential but challenged by manufacturer heterogeneity.	**Strongest**. API specifications, platform rules, and boundary resources constitute the dominant trust architecture (Tiwana, 2013).

The complexity paradox manifests across this spectrum. All smart contracts reduce complexity through standardization, automation, and deterministic execution—eliminating counterparty risk, settlement uncertainties, and enforcement ambiguities (Roeck et al., 2020). Simultaneously, they generate new complexities around code correctness, governance mechanisms, and oracle dependencies (Albizri & Appelbaum, 2021; Goldsby & Hanisch, 2022). The framework's three levels manifest clearly: design-level human control over consensus mechanisms and governance processes shapes trust parameters (Renwick & Gleasure, 2020); operations-level deterministic execution among millions of transactions generates emergent cross-contract interactions; experience-level trust bifurcates between DeFi users conducting elaborate risk assessments informed by audit reports and community signals (Bourveau et al., 2024) and enterprise participants relying on institutional relationships (Goldsby & Hanisch, 2022). The DAO incident illustrates the interplay: a design-level vulnerability produced an operations-level event (the system executing exactly as coded), prompting a design-level intervention (Ethereum's hard fork) that proved controversial precisely because it violated expectations of operational autonomy (DuPont, 2017).

4.2 Artificial Intelligence Systems

AI systems, particularly those employing machine learning, manifest the complexity paradox with particular force: they reduce complexity through pattern recognition exceeding human capacity (Fuster et al., 2022; Obermeyer et al., 2019) while generating epistemic inaccessibility where decision processes remain opaque even with full model access (Rudin, 2019) and emergent drift where models perpetuate biases or fail under distribution shifts (D'Amour et al., 2022). Expectations focus on probabilistic reliability rather than deterministic certainty: a radiologist using diagnostic AI expects statistical accuracy across populations, not certainty in individual cases (Yin et al., 2019). Risk acceptance involves

acknowledging fundamental opacity alongside the possibility of unexplainable or biased outcomes (Kleinberg et al., 2018). Reliant engagement develops through graduated human-AI collaboration, with initial implementations typically augmenting rather than replacing human judgment (Lebovitz et al., 2022).

AI reveals distinctive experience-level dynamics. Trust in virtual and embedded AI often starts high and decreases with experience, inverting the typical human trust trajectory (Glikson & Woolley, 2020). Algorithm aversion compounds this: users reject AI recommendations after observing errors even when AI outperforms humans on average (Burton et al., 2020; Dietvorst et al., 2015, 2018). These dynamics are irreducible to either design-level certification or operations-level performance.

4.3 Internet of Things Systems

IoT systems generate the complexity paradox through the physical-digital interface: individual devices reduce complexity by automating physical interactions, yet aggregating millions of connected devices produces emergent complexities around interoperability, security vulnerabilities, and cross-device coordination (Gubbi et al., 2013; Roman et al., 2013). Drift in IoT thus arises from coupling, updates, and environmental change—amplified by but not dependent on adaptive learning (Benbya et al., 2020).

The triadic components take on a distinctively embodied quality. Expectations encompass physical reliability: autonomous vehicle users trust sensor accuracy, actuator responsiveness, and system resilience to environmental interference (Litman, 2025). Risk acceptance extends to physical safety and cascading failure possibilities where irreversibility of physical actions amplifies consequences (Singh et al., 2020). Reliant engagement develops through embodied experience—trust in smart home systems builds through daily encounters; industrial IoT trust

emerges from observed operational improvements (Boyes et al., 2018). This embodied dimension creates trust dynamics differing qualitatively from purely digital interactions.

4.4 Platform Ecosystems and APIs

Platform ecosystems exhibit lower operational autonomy than AI or smart contracts yet create substantial complexity through interaction effects, ecosystem governance, and heterogeneous participant coordination (de Reuver et al., 2018; Hanisch et al., 2026). Platforms reduce complexity through common interfaces abstracting financial, computational, and informational functions (Ghazawneh & Henfridsson, 2013) while generating complexities through versioning challenges, dependency management, and systemic propagation of changes across interconnected services (Constantinides et al., 2018). Expectations center on API stability and backward compatibility (Trienekens et al., 2004). Risk acceptance involves acknowledging platform dependencies and potential lock-in (Wareham et al., 2014). Most distinctively, reliant engagement becomes architecturally embedded: initial API experiments evolve into comprehensive platform strategies where core processes depend on ecosystem services (Jacobides et al., 2018). This creates switching costs making reliance increasingly consequential over time. Protocol-based interaction structuring dominates the enabling mechanism configuration, with API specifications and platform governance constituting the primary trust-enabling architecture (Hanseth & Lyytinen, 2010; Tiwana, 2013).

4.5 Synthesis

All four contexts exhibit Luhmann's complexity paradox: smart contracts reduce counterparty risk while creating governance complexity, AI simplifies pattern recognition while creating interpretability challenges, IoT automates physical processes while generating coordination complexities, and platforms standardize interactions while creating dependency risks. The triadic structure manifests across all contexts with a consistent dual temporal pattern—initial formation shaped by design-level characteristics, ongoing calibration through operational

experience—yet components adapt: expectations range from deterministic to probabilistic to embodied to contractual; risk acceptance spans code vulnerabilities to physical safety to platform dependencies; reliant engagement varies from irreversible commitment to graduated collaboration to architectural embedding. The three analytical levels prove illuminating across all contexts, with all systems exhibiting clear distinctions between design-level human authority, operations-level algorithmic function, and experience-level trust dynamics. This consistency, combined with sensitivity to contextual variation, reveals digital trust as a family of related phenomena adapted to different algorithmic configurations while sharing the structure our framework identifies.

5 Discussion

5.1 Contributions

This paper offers three contributions to research on trust in algorithmic environments.

5.1.1 Extending Luhmann's Trust Theory to Algorithmic Environments

Our framework extends Luhmann's (1979, 1995) insight that trust serves complexity reduction to a novel context he did not directly theorize: environments where algorithmic systems generate emergent complexity through operations exceeding human comprehension. Luhmann distinguished personal trust (based on familiarity with individuals) from system trust (reliance on abstract functional systems). Algorithmic systems fit neither Luhmann's personal nor system trust categories, creating a gap our concept of operative complexity addresses.

The extension operates along three dimensions. First, we identify *operative complexity*—complexity emerging from autonomous technical operations rather than human communication or institutional procedures—as a distinctive condition requiring trust adaptation. Operative complexity arises when algorithmic systems produce consequential outcomes through processes that are epistemically inaccessible and subject to emergent drift,

creating uncertainty that traditional trust mechanisms cannot fully address. Traditional trust remains operative at the design level, where human actors assess providers, evaluate certifications, and establish governance arrangements. Operative complexity identifies the residual uncertainty that these mechanisms leave unresolved once algorithmic systems begin operating autonomously.

Second, we specify how Luhmann's complexity paradox manifests with distinctive intensity in algorithmic environments. Algorithmic systems amplify the paradox through temporal compression, emergent non-deducibility, and coupling at scale (Section 2.3). The contribution lies in demonstrating, through four empirical contexts, that this amplified paradox proves irreducible across diverse algorithmic configurations—from smart contracts' deterministic execution through AI's probabilistic learning to platforms' ecosystem coordination. Each context exhibits the same underlying dynamic: greater algorithmic sophistication resolves certain complexities while generating qualitatively new forms.

Third, we extend trust theory's treatment of risk. Luhmann emphasized trust as emerging where potential harm accompanies engagement. In algorithmic environments, the risk profile transforms: actors face technical vulnerabilities following heavy-tailed distributions, emergent behaviors no prior analysis could anticipate, and systemic cascades enabled by algorithmic speed and coupling (Pasquale, 2015; Schuetz et al., 2025). Digital trust encompasses these distinctive risk categories, expanding the conditions under which Luhmann's complexity-reduction insight applies.

5.1.2 The Dual Temporal Structure of Digital Trust

Our second contribution theorizes the dual temporal structure through which digital trust forms. Critical trust parameters are established at the design level through architectural decisions—code logic, governance protocols, oversight mechanisms—then continuously

calibrated at the experience level through operational interaction. Prior work established the complementarity between formal governance and relational trust: Poppo and Zenger (2002) demonstrated this for contractual arrangements, and McKnight et al. (1998) distinguished initial trust from trust in ongoing relationships. Our contribution lies in specifying how this dual structure manifests distinctively in algorithmic environments.

The distinctiveness is twofold. First, the operational phase involves a non-human autonomous agent whose behavior drifts from design intent through exogenous, endogenous, and interaction-based pathways (Ciborra, 2002), creating calibration challenges with no precise parallel in human contracting. A smart contract's execution environment changes as other contracts compose with it; a machine learning model's performance shifts as the data distribution diverges from training conditions. The trust object is, in a meaningful sense, a moving target. Second, the three analytical levels operate on different timescales that interact asymmetrically: design changes occur slowly and require substantial coordination; operational patterns shift rapidly as systems interact with dynamic environments; experience-level trust builds gradually but can erode rapidly through negative events or external revelations (Dietvorst et al., 2015; Schmidt et al., 2020). This temporal asymmetry between trust formation and trust erosion extends existing insights about trust's fragility (Slovic, 1993) to environments where the trust object itself undergoes emergent drift.

The multilevel analysis also resolves persistent difficulties in theorizing algorithmic agency. At the design level, human agency dominates; at the operations level, algorithmic systems exhibit operational autonomy; at the experience level, agency becomes distributed as humans and systems jointly shape outcomes (Kiviat, 2019; Korsgaard et al., 2025; Murray et al., 2021). This locates different forms of consequential action at different analytical levels rather than debating whether algorithms "have" agency in a metaphysical sense—a move that

provides IS research with a tractable framework for studying human-algorithmic coordination without requiring resolution of deep ontological questions about machine agency.

5.1.3 Trust Transfer in Algorithmic Environments

Our third contribution addresses trust transfer, which Schuetz et al. (2025) identify as increasingly important yet undertheorized. We specify conditions under which transfer between algorithmic systems becomes more or less likely, and connect these conditions to established trust formation mechanisms.

Transfer likelihood increases when systems share functional characteristics users have learned to associate with reliable operation: consistent interface conventions, transparent performance feedback, and established security protocols. Transfer likelihood decreases when systems exhibit novel characteristics users cannot map to prior experience, when prior negative experiences create wariness toward apparently similar systems, or when contextual factors highlight differences rather than similarities. This conditional specification gives the framework discriminatory power—it can account for cases where trust does not flow between superficially similar systems as well as cases where it does.

The framework connects these dynamics to established mechanisms. Familiarity with one algorithmic system facilitates trust in functionally similar systems, paralleling Gefen's (2000) findings on familiarity-based trust formation. Algorithmic systems embedded in trusted institutional contexts inherit trust from those contexts, extending McKnight et al.'s (1998) analysis of institution-based trust. Söllner et al. (2016) identified trust networks spanning multiple targets; algorithmic systems create new pathways within such networks through shared enabling mechanisms. When actors encounter a novel system exhibiting enabling mechanisms familiar from prior systems—consistent autonomous execution, depersonalized processing, transparent monitoring, protocol-based structuring—transfer operates through

recognition of shared operational characteristics rather than surface similarity or institutional endorsement alone.

This transfer mechanism also illuminates cascading trust failures. When a prominent algorithmic system breaches expectations, trust erosion propagates to other systems sharing similar enabling mechanisms, even when the failure is implementation-specific. As algorithmic systems become increasingly interconnected through shared protocols, APIs, and data flows, understanding these transfer dynamics becomes essential for explaining both rapid trust formation in novel systems and the systemic fragility that interconnection creates.

5.2 Practical Implications

The framework yields implications for organizations, policymakers, and system designers.

For organizations, digital trust as a response to algorithmically-generated complexity means trust-building strategies must address both complexity-reducing and complexity-generating aspects of algorithmic deployment. Pursuing transparency as a singular strategy is insufficient; organizations should combine design-level governance, operations-level monitoring, and experience-level communication to help stakeholders navigate irreducible uncertainty (Curchod et al., 2020; Glikson & Woolley, 2020). Building organizational capacity for algorithmic engagement requires competencies spanning algorithmic literacy, risk assessment, and ecosystem awareness (Kane, 2019). The multilevel framework indicates that trust strategies must coordinate across design, operations, and experience—requiring structures that bridge technical and business functions and anticipate how trust requirements shift as algorithmic systems are progressively embedded in organizational processes.

For policymakers, regulatory approaches must account for the complexity paradox. Mandating transparency addresses design-level concerns but cannot eliminate the emergent drift and epistemic inaccessibility that arise at the operations level. Effective regulation

requires adaptive governance frameworks accommodating algorithmic change rather than static compliance regimes (Danaher et al., 2017; Kaminski, 2018). The EU's AI Act addresses design-level requirements but faces challenges in governing operations-level dynamics where emergent behaviors exceed any pre-deployment assessment.

For system designers, the four enabling mechanisms provide concrete guidance. Consistent autonomous execution, depersonalized processing, transparent performance monitoring, and protocol-based interaction structuring each contribute to trust formation, but none alone suffices. Design should attend to how mechanisms combine and to how design-level choices create conditions within which operations-level trust dynamics unfold (Friedman & Hendry, 2019; Shneiderman, 2022).

5.3 Limitations and Future Research

Several limitations frame the contribution and open avenues for inquiry.

First, the framework's conceptual nature requires empirical validation. The triadic structure and three analytical levels provide analytical tools whose empirical utility remains to be demonstrated. Measurement approaches capturing digital trust's integrated structure rather than isolated components (Gefen et al., 2003; Venkatesh et al., 2016), and mixed-method designs combining computational analysis of algorithmic behavior with qualitative investigation of trust dynamics, represent productive directions.

Second, the framework presumes algorithmic systems operating with sufficient autonomy to generate emergent complexity. The boundary between systems requiring digital trust and those manageable through traditional mechanisms is a gradient, and empirical research is needed to identify where along this gradient the framework's constructs become most analytically productive (Rai et al., 2019).

Third, the framework draws on complexity theory for its foundational insight—trust as a mechanism for reducing otherwise unmanageable complexity—while developing constructs (epistemic inaccessibility, emergent drift) that are IS-theoretic rather than complexity-theoretic in origin. Complex social, biological, and physical systems exhibit emergence and unpredictability of their own (Miller & Page, 2007). What distinguishes the algorithmic case is the joint convergence of three conditions no physical or social complex system exhibits together. The first is the gap between promised and operational inspectability: algorithmic systems are constructed from formal specifications and are increasingly subject to audit and documentation mandates, yet at operational scale the inspectable trace exceeds human cognitive reconstruction (Burrell, 2016; Humphreys, 2009). Bee swarms and sandpiles are similarly inscrutable in the aggregate but occupy no trustee position. The second is endogenously structured drift: algorithmic systems change through intentional patching, retraining, and versioning (Ciborra, 2002; Pentland et al., 2020), so the trustee a user formed expectations about may not be the trustee operating subsequently—a designed instability absent from spontaneous emergence. The third is the delegation relationship itself: trust in algorithmic systems is a principal–agent problem (Baird & Maruping, 2021) in which the trustee receives delegated authority and the trustor accepts vulnerability accordingly. Other complex systems may exhibit one or another of these properties; only algorithmic systems combine all three in the trustee position. Deeper integration of complexity-theoretic and systems-theoretic perspectives on trust by extending the framework to complex digital systems beyond the algorithmic case is a productive direction for future research.

Fourth, the systems-theoretic foundation does not directly address questions of power, justice, and distribution. Critical perspectives on algorithmic governance (Kellogg et al., 2020; Zuboff, 2019) complement the framework by illuminating whose trust matters, who bears the costs of algorithmic complexity, and how trust formation intersects with structural inequality.

Integrating systemic and critical perspectives could enrich understanding of digital trust's distributional consequences.

Fifth, the framework presupposes that actors maintain meaningful choice about engagement. In contexts of algorithmic monopoly or regulatory mandate, frameworks emphasizing power, compliance, or resistance prove more illuminating (Kellogg et al., 2020; Möllering, 2006).

Finally, the relationship between digital trust and organizational learning deserves investigation. As organizations accumulate experience with algorithmic systems, how do collective trust dynamics change? Process-oriented methodologies, including ethnographic studies examining how design, operations, and experience levels mutually constitute each other through organizational practice (Leonardi, 2011; Orlikowski & Scott, 2008), could reveal temporal dynamics the conceptual framework identifies but cannot empirically trace.

6 Conclusion

Algorithmic systems are reshaping organizational coordination, creating emergent complexity that traditional trust mechanisms cannot fully address. Our framework specifies how digital trust emerges as a systemic response to Luhmann's (1979, 1995) complexity paradox—systems deployed to reduce complexity simultaneously generating new uncertainty—a paradox amplified in algorithmic environments through temporal compression, emergent non-deducibility, and coupling at scale. By identifying operative complexity as the distinctive condition, theorizing the dual temporal structure through which digital trust forms across design, operations, and experience levels, and specifying the conditions under which trust transfers through shared enabling mechanisms, the framework provides analytical tools for understanding how organizations navigate irreducible algorithmic uncertainty. As algorithmic systems assume increasingly consequential organizational functions, understanding how trust operates under conditions of epistemic inaccessibility and emergent drift becomes essential—

not only for theory but for the organizational practice of delegating to systems whose full operational behavior cannot be known in advance.

References

Albizri, A., & Appelbaum, D. (2021). Trust but Verify: The Oracle Paradox of Blockchain Smart Contracts. *Journal of Information Systems*, *ISYS-19-024*. https://doi.org/10.2308/ISYS-19-024

Alter, S. (2008). Defining information systems as work systems: Implications for the IS field. *European Journal of Information Systems*, *17*(5), 448–469. https://doi.org/10.1057/ejis.2008.37

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Arifovic, J., He, X., & Wei, L. (2022). Machine learning and speed in high-frequency trading. *Journal of Economic Dynamics and Control*, *139*, 104438. https://doi.org/10.1016/j.jedc.2022.104438

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, *45*(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Barile, S., Simone, C., Iandolo, F., & Laudando, A. (2022). Platform-based innovation ecosystems: Entering new markets through holographic strategies. *Industrial Marketing Management*, *105*, 467–477. https://doi.org/10.1016/j.indmarman.2022.07.003

Beck, R., Czepluch, J. S., Lollike, N., & Malone, S. (2016). *Blockchain–the gateway to trust-free cryptographic transactions*. 1–14.

Beck, R., Müller-Bloch, C., & King, J. L. (2018). Governance in the blockchain economy: A framework and research agenda. *Journal of the Association for Information Systems*, *19*(10), 1.

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 4.

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS Quarterly*, *44*(1), 1–17. https://doi.org/0.25300/MISQ/2020/13304

Bourveau, T., Brendel, J., & Schoenfeld, J. (2024). Decentralized Finance (DeFi) assurance: Early evidence. *Review of Accounting Studies*, *29*(3), 2209–2253. https://doi.org/10.1007/s11142-024-09834-8

Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in Industry*, *101*, 1–12.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512.

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239.

Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*, *40*(5), 685–706. https://doi.org/10.1016/j.emj.2022.06.001

Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology*, *33*(4), 685–703.

Chen, L., Tong, T. W., Tang, S., & Han, N. (2022). Governance and Design of Digital Platforms: A Review and Future Research Directions on a Meta-Organization. *Journal of Management*, *48*(1), 147–184. https://doi.org/10.1177/01492063211045023

Chen, R. R., Chen, K., & Ou, C. X. J. (2023). Facilitating interorganizational trust in strategic alliances by leveraging blockchain-based systems: Case studies of two eastern banks. *International Journal of Information Management*, *68*, 102521. https://doi.org/10.1016/j.ijinfomgt.2022.102521

Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, *13*. https://doi.org/10.1016/j.jbvi.2019.e00151

Ciborra, C. (2002). *The labyrinths of information: Challenging the wisdom of systems: Challenging the wisdom of systems*. OUP Oxford.

Coleman, J. S. (1990). *Foundations of Social Theory*. Belknap Press of Harvard University Press.

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and Infrastructures in the Digital Age. *Information Systems Research*, *29*(2), 381–400. https://doi.org/10.1287/isre.2018.0794

Corning, P. A. (2002). The re-emergence of "emergence": A venerable concept in search of a theory. *Complexity*, *7*(6), 18–30. https://doi.org/10.1002/cplx.10043

Cremer, J., Konstantelos, I., & Strbac, G. (2019). From Optimization-Based Machine Learning to Interpretable Security Rules for Operation. *IEEE Transactions on Power Systems*, *34*, 3826–3836. https://doi.org/10.1109/TPWRS.2019.2911598

Cui, M., Li, W., & Kamoche, K. (2026). Building Trust in Decision-Support Artificial Intelligence: A Boundary Spanning Perspective. *Information Systems Journal*, *36*(3), 435–456. https://doi.org/10.1111/isj.70015

Curchod, C., Patriotta, G., Cohen, L., & Neysen, N. (2020). Working for an algorithm: Power asymmetries and agency in online work settings. *Administrative Science Quarterly*, *65*(3), 644–676.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., & Hoffman, M. D. (2022). Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, *23*(226), 1–61.

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., & Morison, J. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, *4*(2), 2053951717726554.

De Filippi, P., Mannan, M., & Reijers, W. (2020). Blockchain as a confidence machine: The problem of trust & challenges of governance. *Technology in Society*, *62*. https://doi.org/10.1016/j.techsoc.2020.101284

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*(2), 124–135.

Demetis, D. S., & Lee, A. S. (2018). When Humans Using the IT Artifact Becomes IT Using the Human Artifact. *Journal of the Association for Information Systems*, 929–952. https://doi.org/10.17705/1jais.00514

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

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

DuPont, Q. (2017). Experiments in algorithmic governance: A history and ethnography of "The DAO," a failed decentralized autonomous organization. In *Bitcoin and beyond* (pp. 157–177). Routledge.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*.

Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, *15*(1), 9–29.

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Mit Press.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, *77*(1), 5–47.

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega*, *28*(6), 725–737.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 51–90.

Gefen, D., Wyss, S., & Lichtenstein, Y. (2008). Business Familiarity as Risk Mitigation in Software Development Outsourcing Contracts1. *Management Information Systems Quarterly*, *32*(3), 531–551. https://doi.org/10.2307/25148855

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *Information Systems Journal*, *23*(2), 173–192.

Giddens, A. (1990). *The consequences of modernity* (Reprint). Polity Press.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/doi.org/10.5465/annals.2018.0057

Goldsby, C., & Hanisch, M. (2022). The Boon and Bane of Blockchain: Getting the Governance Right. *California Management Review*, *64*(3), 141–168. https://doi.org/10.1177/00081256221080747

Gregory, R. W., Beck, R., Henfridsson, O., & Yaraghi, N. (2024). Cooperation Among Strangers: Algorithmic Enforcement of Reciprocal Exchange with Blockchain-Based Smart Contracts. *Academy of Management Review*, amr.2023.0023. https://doi.org/10.5465/amr.2023.0023

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660.

Hacker, J., Miscione, G., Felder, T., & Schwabe, G. (2023). Commit or not? How blockchain consortia form and develop. *California Management Review*, *65*(3), 110–131. https://doi.org/10.1177/00081256231175530

Halaburda, H., Levina, N., & Min, S. (2024). Digitization of Transaction Terms within TCE: Strong Smart Contract as a New Mode of Transaction Governance. *MIS Quarterly*, *48*(2), 825–846. https://doi.org/10.25300/MISQ/2023/17818

Hanisch, M., Roozen, P., & Theodosiadis, V. (2026). Success and Failure in Blockchain-Based Interorganizational Ecosystems: A Governance Perspective. *Journal of Management*, 01492063261420752. https://doi.org/10.1177/01492063261420752

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building internet. *Journal of Information Technology*, *25*(1), 1–19.

Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. *Electronic Commerce Research and Applications*, *29*, 50–63. https://doi.org/10.1016/j.elerap.2018.03.005

Hawlitschek, F., Notheisen, B., & Teubner, T. (2020). A 2020 perspective on "The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy". *Electronic Commerce Research and Applications*, *40*. https://doi.org/10.1016/j.elerap.2020.100935

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Jackson, S., & Panteli, N. (2023). Trust or mistrust in algorithmic grading? An embedded agency perspective. *International Journal of Information Management*, *69*, 102555. https://doi.org/10.1016/j.ijinfomgt.2022.102555

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*(8), 2255–2276.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). *Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI*. 624–635.

Kaminski, M. E. (2018). Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *S. Cal. L. Rev.*, *92*, 1529.

Kane, G. (2019). The technology fallacy: People are the real key to digital transformation. *Research-Technology Management*, *62*(6), 44–49.

Kellogg, K., Valentine, M., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/doi.org/10.5465/annals.2018.0174

Killoran, J., Park, A., & Kietzmann, J. (2024). Delineation of Reasoning, Intentional Disclosure, and the Potential for Harm: An Extension of Vanneste and Puranam's "Artificial Intelligence, Trust, and Perceptions of Agency". *Academy of Management Review*, amr.2024.0193. https://doi.org/10.5465/amr.2024.0193

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, *72*(3), 967–998.

Kiviat, B. (2019). The moral limits of predictive practices: The case of credit-based insurance scores. *American Sociological Review*, *84*(6), 1134–1158.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, *10*, 113–174.

Koghut, M., Al-Tabbaa, O., Lee, S. H., & Meyer, M. (2021). A Blockchain-based inter-organisational relationships: Social and innovation implications. *Academy of Management Proceedings*, *2021*(1), 14553. https://doi.org/10.5465/AMBPP.2021.14553abstract

Koh, T. K., Fichman, M., & Kraut, R. E. (2012). Trust Across Borders: Buyer-Supplier Trust in Global Business-to-Business E-Commerce. *Journal of the Association for Information Systems*, *13*(11), 886–922.

Komiak, S. Y., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly*, 941–960.

Korsgaard, M. A., Cooper, C. D., Mayer, K. J., Poppo, L., & Zaheer, A. (2025). The Boundaries of Trust in a New Era. *Academy of Management Review*, 1–11. https://doi.org/10.5465/amr.2024.0605

Lacity, M. C., Schuetz, S. W., Kuai, L., & Steelman, Z. R. (2025). IT's a matter of trust: Literature reviews and analyses of human trust in information technology. *Journal of Information Technology*, *40*(3), 286–313. https://doi.org/10.1177/02683962231226397

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473.

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1), 126–148.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, *35*(1), 147–167.

Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2017). A survey on the security of blockchain systems. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2017.08.020

Litman, T. (2025). *Autonomous vehicle implementation predictions: Implications for transport planning* (p. 49). Victoria Transport Policy Institute Victoria, BC, Canada. https://www.vtpi.org/avip.pdf

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

Luhmann, N. (1979). *Trust and Power: Two Works by Niklas Luhmann. Translation of German originals Vertrauen [1968] and Macht [1975]*. John Wiley.

Luhmann, N. (1988). Familiarity, Confidence, Trust: Problems and Alternatives. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 94–107). Basil Blackwell. http://www.sociology.ox.ac.uk/papers/luhmann94-107.pdf

Luhmann, N. (1995). *Social Systems*. Stanford University Press.

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, *32*(4), 1993–2020. https://doi.org/10.1007/s12525-022-00605-4

Lumineau, F., Wang, W., & Schilke, O. (2021). Blockchain Governance—A New Way of Organizing Collaborations? *Organization Science*, *32*(2), 500–521. https://doi.org/10.1287/orsc.2020.1379

Lumineau, F., Wang, W., & Schilke, O. (2025). Reframing Blockchain's Promise: A Commentary on Gregory, Beck, Henfridsson, and Yaraghi's "Cooperation among Strangers". *Academy of Management Review*, amr.2024.0375. https://doi.org/10.5465/amr.2024.0375

MacKenzie, D. (2019). How algorithms interact: Goffman's 'interaction order'in automated trading. *Theory, Culture & Society*, *36*(2), 39–59.

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

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a Specific Technology: An Investigation of Its Components and Measures. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, *23*(3), 473–490.

Miller, J. H., & Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life (Princeton Studies in Complexity)*. Princeton university press.

Möllering, G. (2006). *Trust: Reason, Routine, Reflexivity*. Emerald Group Publishing.

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Obermeier, D., & Henkel, J. (2025). Deductive Certainty? Exploring the Boundaries of Trust Formation in Smart Contracts on Blockchains. *MIS Quarterly*, *49*(4), 1513–1538.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: Challenging the Separation of Technology, Work and Organization. *Academy of Management Annals*, *2*(1), 433–474.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. https://books.google.co.uk/books?id=TumaBQAAQBAJ

Pavlou, P. A., & Fygenson, M. (2006). Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior1. *Management Information Systems Quarterly*, *30*(1), 115–143. https://doi.org/10.2307/25148720

Pentland, B. T., Liu, P., Kremser, W., & Hærem, T. (2020). The Dynamics of Drift in Digitized Processes. *MIS Quarterly*, *44*(1), 19–48. JSTOR.

Poppo, L., & Zenger, T. R. (2002). Do formal contracts and relational governance function as substitutes or complements? *Strategic Management Journal*, *23*(8), 707–725.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14.

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

Ramaul, L., Ritala, P., Kostis, A., & Aaltonen, P. (2026). Rethinking How We Theorize AI in Organization and Management: A Problematizing Review of Rationality and Anthropomorphism. *Journal of Management Studies*, *63*(2), 761–807. https://doi.org/10.1111/joms.13246

Renwick, R., & Gleasure, R. (2020). Those who control the code control the rules: How different perspectives of privacy are being written into the code of blockchain systems. *Journal of Information Technology*, 026839622094440. https://doi.org/10.1177/0268396220944406

Roeck, D., Sternberg, H., & Hofmann, E. (2020). Distributed ledger technology in supply chains: A transaction cost perspective. *International Journal of Production Research*, 1–18. https://doi.org/10.1080/00207543.2019.1657247

Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, *57*(10), 2266–2279.

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

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.

Sadr, H., Nazari, M., Khodaverdian, Z., Farzan, R., Yousefzadeh-Chabok, S., Ashoobi, M. T., Hemmati, H., Hendi, A., Ashraf, A., & Pedram, M. M. (2025). Unveiling the potential of artificial intelligence in revolutionizing disease diagnosis and prediction: A comprehensive review of machine learning and deep learning approaches. *European Journal of Medical Research*, *30*(1), 418.

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, *29*(4), 260–278.

Schuetz, S., Kuai, L., Lacity, M. C., & Steelman, Z. (2025). A qualitative systematic review of trust in technology. *Journal of Information Technology*, *40*(1), 55–76. https://doi.org/10.1177/02683962241254392

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, 59–68. https://doi.org/10.1145/3287560.3287598

Shaikh, M., & Vaast, E. (2023). Algorithmic interactions in open source work. *Information Systems Research*, *34*(2), 744–765.

Shapiro, S. P. (1987). The Social Control of Impersonal Trust. *American Journal of Sociology*, *93*(3), 623–658.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551.

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Singh, D., Pati, B., Panigrahi, C. R., & Swagatika, S. (2020). Security Issues in IoT and their Countermeasures in Smart City Applications. In B. Pati, C. R. Panigrahi, R. Buyya, & K.-C. Li (Eds), *Advanced Computing and Intelligent Engineering* (pp. 301–313). Springer Singapore.

Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, *13*(6), 675–682.

Söllner, M., Hoffmann ,Axel, & and Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, *25*(3), 274–287. https://doi.org/10.1057/ejis.2015.17

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd edn). Cambridge University Press.

Taleb, N. N. (2010). *The Black Swan: The Impact of the Highly Improbable: With a new section:" On Robustness and Fragility"* (Vol. 2). Random house trade paperbacks.

Tiwana, A. (2013). *Platform ecosystems: Aligning architecture, governance, and strategy*. Newnes.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research*, *21*(4), 675–687. https://doi.org/10.1287/isre.1100.0323

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & Van Moorsel, A. (2020). *The relationship between trust in AI and trustworthy machine learning technologies*. 272–283.

Trienekens, J., Bouman, J., & Zwan, M. (2004). Specification of Service Level Agreements: Problems, Principles and Practices. *Software Quality Journal*, *12*, 43–57. https://doi.org/10.1023/B:SQJO.0000013358.61395.96

Utz, M., Johanning, S., Roth, T., Bruckner, T., & Strüker, J. (2023). From ambivalence to trust: Using blockchain in customer loyalty programs. *International Journal of Information Management*, *68*, 102496. https://doi.org/10.1016/j.ijinfomgt.2022.102496

Vance, A., Elie-Dit-Cosaque, C., & Straub, D. W. (2008). Examining Trust in Information Technology Artifacts: The Effects of System Quality and Culture. *Journal of*

Management Information Systems, *24*(4), 73–100. https://doi.org/10.2753/MIS0742-1222240403

Vanneste, B. S., & Puranam, P. (2024). Artificial Intelligence, Trust, and Perceptions of Agency. *Academy of Management Review*, amr.2022.0041. https://doi.org/10.5465/amr.2022.0041

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112.

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, *17*(5), 328–376.

Vidan, G., & Lehdonvirta, V. (2019). Mine the gap: Bitcoin and the maintenance of trustlessness. *New Media & Society*, *21*(1), 42–59.

Völter, F., Urbach, N., & Padget, J. (2023). Trusting the trust machine: Evaluating trust signals of blockchain applications. *International Journal of Information Management*, *68*, 102429. https://doi.org/10.1016/j.ijinfomgt.2021.102429

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246.

Wareham, J., Fox, P. B., & Cano Giner, J. L. (2014). Technology ecosystem governance. *Organization Science*, *25*(4), 1195–1215.

Weaver, W. (1948). Science and complexity. *American Scientist*, *36*(4), 536–544. JSTOR.

Whetten, D. A. (1989). What Constitutes a Theoretical Contribution? *Academy of Management Review*, *14*(4), 490–495.

Yin, M., Wortman Vaughan, J., & Wallach, H. (2019). *Understanding the effect of accuracy on trust in machine learning models*. 1–12.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Algorithmic decision-making and the control problem. *Minds and Machines*, *29*(4), 555–578.

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020). *Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making*. 295–305.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

Zucker, L. G. (1986). Production of Trust: Institutional Sources of Economic Structure, 1840-1920. In B. M. Staw & L. L. Cummings (Eds), *Research in Organizational Behavior* (Vol. 8, pp. 53–111). JAI Press.