Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.085515

REVIEW

Trust and Cybersecurity Behaviours: A Scoping Review

Shadi Melebari, Muhammad Atif Ur Rehman*, Ali Kashif Bashir and Mohammed Al-Khalidi

Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK

*Corresponding Author: Muhammad Atif Ur Rehman. Email: m.atif.ur.rehman@mmu.ac.uk

Received: 12 May 2026; Accepted: 20 July 2026; Published: 15 September 2026

ABSTRACT: Trust plays an important role in shaping human behaviour in cybersecurity contexts, influencing how individuals interact with digital systems and respond to security practices. However, existing research on trust and cybersecurity behaviours remains fragmented, with different studies adopting varied definitions of trust and examining a wide range of behavioural outcomes. This scoping review aims to systematically map and synthesise the literature on trust and cybersecurity behaviours. Following the PRISMA-ScR guidelines, relevant studies were identified, screened, and analysed to examine how trust has been conceptualised, what types of behaviour have been studied, and which theoretical and methodological approaches have been used. The findings show that trust is a multi-dimensional construct, including system, institutional, interpersonal, and generalised forms, and that it has been examined in relation to different cybersecurity behaviours across human and system-level contexts. However, the direction of the trust–behaviour relationship remains unclear or mixed in the majority of included studies, particularly within the human-focused subset of 19 studies, which suggests that current evidence is insufficient to establish a stable or consistent relationship between trust and cybersecurity behaviour. In the human-focused subset, adoption and compliance behaviours were the most commonly examined, while social engineering received less attention; authentication and technical risk behaviours were not examined in any human-focused study. This review contributes to the literature by providing a structured overview of how trust has been conceptualised, measured, and studied in cybersecurity contexts, and by identifying key conceptual, methodological, and empirical gaps that warrant further investigation.

KEYWORDS: Trust; cybersecurity; risky behaviour; human factors; socio-technical systems; PRISMA-ScR; scoping review

1 Introduction

The rapid digital transformation of modern societies has made cybersecurity a central concern for individuals, organisations, and governments [1]. As reliance on online services and intelligent systems increases, trust plays a central role in shaping users' decisions and their willingness to rely on automated systems [2]. Appropriate trust enables cooperation, compliance, and adoption of technology, while misplaced or excessive trust can lead to overconfidence, negligence, and risky cybersecurity behaviours such as ignoring warnings, sharing credentials, or falling for social-engineering attacks [3]. Understanding how trust influences these behaviours is therefore critical for developing effective human-centred cybersecurity strategies.

Trust may influence different types of security-related decision-making. In cybersecurity, trust can operate as a double-edged factor: misplaced or manipulated trust may increase risk by reducing perceived vulnerability and making users more susceptible to deception and social-engineering attacks [4,5], while appropriately calibrated trust may support compliance with security guidance and engagement with protective practices [6].

Cybersecurity research has traditionally been framed as a technological challenge, with an emphasis on technical defences and system vulnerabilities. However, recent human-centred cybersecurity research increasingly recognises that human factors, including trust, attitudes, privacy and risk perceptions, cognitive biases, and organisational culture, play a critical role in shaping cybersecurity vulnerabilities, resilience, and compliance [1,7].

Trust in this context is multi-layered. At a general level, trust can be understood as a positive expectation that involves some degree of risk or vulnerability [8]. In cybersecurity contexts, this broader concept can appear as system trust, institutional trust, interpersonal trust, or generalised trust. System trust refers to confidence in technology and its reliability [2]. Institutional trust relates to belief in the competence and integrity of organisations and authorities, while interpersonal trust concerns confidence in other individuals, such as colleagues or online peers [9,10]. Generalised trust reflects a broader tendency to extend trust beyond immediate or familiar relationships [8]. These different dimensions of trust may influence behaviour in various ways, depending on social, cultural, organisational, and technical contexts.

Regardless of increasing scholarly attention, research on trust and cybersecurity behaviour remains fragmented. Existing studies adopt diverse theoretical frameworks, ranging from psychological models, including the Theory of Planned Behaviour (TPB) [11], Protection Motivation Theory (PMT) [12], and the Technology Acceptance Model (TAM) [13], to broader socio-technical and technology acceptance approaches. Many findings remain descriptive or inconsistent, and the lack of integrated evidence makes it difficult to determine when trust acts as a protective factor and when it increases risk exposure. Moreover, the evidence base remains methodologically and contextually uneven, with variation in study designs, measurement approaches, populations, and regional coverage.

Given this conceptual diversity, methodological imbalance, and inconsistency in definitions, a scoping review is a suitable approach to mapping existing evidence, clarifying how trust has been operationalised, and identifying gaps to guide future empirical work. Therefore, this review aims to synthesise how trust is conceptualised, measured, and linked to cybersecurity behaviours across different populations, sectors, and contexts.

This review was guided by the following research questions:

1. RQ1: How has trust been defined and operationalised in studies addressing cybersecurity behaviour?
2. RQ2: What types of cybersecurity behaviours are associated with trust in the existing literature?
3. RQ3: What theories and methodologies have been used to evaluate the relationship between trust and cybersecurity behaviour?
4. RQ4: What conceptual, empirical, and methodological gaps remain in this research domain?

It is important to note that the reviewed literature spans both human-centred studies involving adult participants and system-level, machine-learning, or testbed-based studies in which trust is operationalised as a technical property. Both strands are included because they reflect how trust has been studied in cybersecurity research. Where findings depend on participant characteristics or human-focused behaviour, the analysis is reported separately for the human-only subset (N = 19) to avoid conflating human-centred and system-level evidence.

By addressing these questions, the review contributes to a clearer understanding of trust as a socio-technical construct within cybersecurity and establishes a roadmap for future research in this growing field.

Paper Motivation and Contribution

Although trust is widely recognised as an important factor in cybersecurity, existing research provides limited clarity on how different forms of trust influence cybersecurity behaviours. The definition and measurement of trust in prior studies vary considerably, often leading to mixed or context-dependent findings. As a result, the conditions under which trust supports or undermines cybersecurity behaviour remain poorly understood.

Several limitations in the current literature motivate this review. Research has predominantly focused on system and institutional trust, with far less attention given to interpersonal and generalised trust. Methodologically, the evidence base remains uneven, with variation across human-centred, system-level, cross-sectional, qualitative, and mixed methods. In addition, many studies lack strong theoretical grounding, and researchers have underexplored important sectors, cultures, populations, and emerging technological settings.

To address these issues, this paper presents a scoping review that maps how trust has been conceptualised, operationalised, and linked to cybersecurity behaviours. A scoping approach is well suited to this purpose because it accommodates conceptual diversity and methodological variation while offering a structured overview of the field [14].

The broad conceptual scope of this review, spanning system, institutional, interpersonal, and generalised trust across both human-centred and system-level cybersecurity contexts, is a deliberate methodological choice rather than a limitation. Scoping reviews are specifically designed to map diverse and evolving bodies of literature without imposing narrow domain boundaries [14,15]. Narrowing the scope to a single trust dimension or cybersecurity domain would risk obscuring the very fragmentation and inconsistency that motivates this review. At the same time, clear boundaries were applied: studies focused solely on cryptographic, protocol-level, or mathematical trust mechanisms without behavioural or organisational relevance were excluded, and the review does not claim that different forms of trust operate equivalently across contexts.

This review makes four key contributions to the existing literature. First, it offers a structured overview of trust dimensions used in cybersecurity research, including system, institutional, interpersonal, and generalised trust, bringing these together in a single analytical framework. Second, it maps the types of cybersecurity behaviour examined in relation to trust, covering adoption-compliance, social engineering, authentication, technical risk, and other unclear behavioural categories. Third, it examines the theoretical and methodological approaches applied in existing studies, highlighting key gaps and imbalances that prior surveys have not systematically addressed. Finally, it identifies priority areas for future research, including the need for stronger theoretical integration, more robust methods, and greater attention to interpersonal trust, emerging technologies, and diverse cultural contexts.

The rest of this paper is organised as follows: Section 2 (Review of Related Surveys) positions the present study within the existing survey literature and reviews research on cybersecurity behaviour, human factors, and trust. Section 3 describes the methodology and search strategy used in this scoping review. Section 4 presents the results, including descriptive patterns and thematic findings. Section 5 discusses the implications of these results and introduces a conceptual framework linking trust and cybersecurity behaviours. Section 6 outlines limitations and directions for future research, and Section 7 concludes the paper.

2 Review of Related Surveys

A growing number of survey and review papers have examined human aspects of cybersecurity, including cybersecurity behaviours, human factors, and trust in digital systems. However, these works vary considerably in their scope, conceptual framing, and analytical focus. Rather than synthesising empirical findings, this section aims to position the present scoping review within the existing survey-level literature and clarify how prior surveys have approached cybersecurity behaviour and trust, as well as where key conceptual gaps remain. Existing survey-level work is summarised in Table 1, highlighting that trust is rarely treated as a central analytical construct.

2.1 Surveys on Cybersecurity Behaviours

Several surveys focus primarily on cybersecurity behaviour, such as compliance with security policies, non-compliance, risky decision-making, and security awareness. These reviews provide valuable overviews of behavioural drivers, organisational influences, and individual factors that shape how users engage with security practices in both personal and professional contexts. Some emphasise compliance with information security policies, while others examine how cybersecurity behaviours are defined and measured across studies.

Despite their relevance, these surveys generally do not treat trust as a central analytical construct. Trust is either absent or discussed only implicitly, for example, as a background assumption related to technology use or organisational structures. As a result, while these reviews offer important insights into cybersecurity behaviours more broadly, they do not systematically examine how trust influences cybersecurity behaviours across organisational and everyday contexts [16–18].

2.2 Surveys on Human Factors and Socio-Technical Perspectives

A second group of surveys adopts a broader human factors or socio-technical perspective, addressing user awareness, social engineering, organisational processes, and the interaction between people and security technologies. These reviews highlight the importance of human vulnerabilities and decision-making in cybersecurity incidents, including susceptibility to social engineering and manipulation, and they often argue for more human-centred security approaches [19,20].

2.3 Surveys Addressing Trust in Digital or Security Contexts

Existing surveys that explicitly address trust in technology or security systems typically approach trust from technical, architectural, or governance-oriented perspectives. These works clarify trust as a socio-technical concept by focusing on system reliability, security mechanisms, or institutional arrangements. However, they generally do not examine how trust relates to specific cybersecurity behaviours, such as risk-taking, non-compliance, or security neglect among users.

Within this literature, trust is sometimes discussed implicitly, particularly in relation to reliance on systems, awareness, or susceptibility to manipulation; however, it is rarely defined, operationalised, or used as an organising lens for reviewing cybersecurity behaviours. As a result, these surveys provide valuable contextual insight into human involvement in cybersecurity but stop short of systematically mapping trust constructs to specific risky cybersecurity behaviours [1,3].

Taken together, these strands of survey-level literature indicate that cybersecurity behaviour, human factors, and trust have largely been reviewed in parallel rather than through an integrated analytical lens that links trust to specific, risky cybersecurity behaviours. Existing surveys either focus on cybersecurity

behaviours without systematically addressing trust or examine trust without mapping its relationship to risky cybersecurity behaviours across different user contexts [21].

Table 1: Comparison of existing survey and review papers related to trust and cybersecurity behaviours.

Author (Year)	Review Type	Main Focus	Cybersecurity Behaviour	Treatment of Trust	Population /Context	Key Scope Limitation
Angraini et al. [18]	Systematic	Information security policy compliance	Compliance and non-compliance	Not addressed	Organisational users	Focuses on policy-compliance determinants without examining trust as a behavioural driver.
Ali et al. [16]	Systematic	Non-compliance to compliance	Information security compliance behaviour	Not addressed	Organisational settings	Examines compliance processes but does not consider trust-related perceptions or beliefs.
Rahman et al. [1]	Scoping	Human factors in cybersecurity	Broad human-centric security behaviours	Implicit	Socio-technical systems	Addresses human factors broadly; trust is not defined or analysed as a construct.
Khando et al. [3]	Systematic	Methods and factors for enhancing information security awareness	Awareness-related, compliant, and risky behaviours	Not addressed	Private/public organisations	Focuses on awareness methods and influencing factors rather than trust as a determinant of cybersecurity behaviour.
Almansoori et al. [17]	Systematic	Cybersecurity behaviour studies and theories	Risky and protective behaviours	Not addressed	Multiple user contexts	Provides a broad behavioural and theoretical landscape but does not examine trust–behaviour relationships.
Kannelønning and Katsikas [21]	Systematic	Assessment and measurement of cybersecurity behaviour	Behavioural measurement and evaluation	Not addressed	Organisational contexts	Concentrates on measurement instruments rather than behavioural drivers such as trust.
This review (2026)	Scoping	Trust and cybersecurity behaviour	Risky and protective cybersecurity behaviours	Explicit multi-dimensional treatment	Adult users and system/ML/ testbed contexts	Systematically maps trust as a central construct across both human-centred and system-level cybersecurity contexts; the human-focused subset (N = 19) is treated separately throughout the analysis.

2.4 Positioning of the Present Scoping Review

As summarised in Table 1, existing surveys have examined cybersecurity behaviours, human factors, and trust from different perspectives. However, these areas have mostly been reviewed separately rather than brought together through a common analytical lens. In many reviews, trust is either not addressed directly or is discussed only as a background factor, without being clearly defined or linked to specific cybersecurity behaviours. Although previous socio-technical and trust-focused studies have considered aspects of this relationship, existing survey-level work has not provided a structured mapping of how trust has been conceptualised, operationalised, and examined across both human-centred and system-level cybersecurity

contexts. This scoping review addresses this gap by bringing together behavioural, socio-technical, and trust-focused literature to provide a clearer overview of how trust has been studied in relation to cybersecurity behaviours. Rather than claiming that no prior work has considered this topic, the review offers a more organised and comprehensive mapping of trust dimensions, behavioural categories, theoretical foundations, and methodological approaches across diverse contexts.

3 Methodology

The methodology for this scoping review was guided by the PRISMA-ScR standards introduced by [22]. The completed PRISMA-ScR checklist is provided in Supplementary Table S1. The aim was to map the existing evidence on how trust has been studied in relation to cybersecurity behaviours among adults, identify key patterns, and highlight areas that require further investigation.

A scoping review design was chosen as it enables a structured overview of a diverse and evolving body of literature, rather than assessing study quality or estimating effect sizes. This approach is particularly appropriate for topics characterised by conceptual variation and methodological diversity.

Studies focusing on system-level, machine learning, or testbed-based trust were included if they contributed to understanding how trust is conceptualised, operationalised, or used to evaluate cybersecurity-related risk, behaviour, or system response. However, studies focused only on purely cryptographic, protocol-level, or mathematical trust mechanisms were excluded if they lacked behavioural, organisational, or cybersecurity risk-related relevance. This allowed the review to capture both human-centred and system-oriented perspectives on trust within cybersecurity research.

3.1 Study Design and Protocol

The review followed the five stages proposed by [14] and further refined by [15]:

1. Identifying the research question.
2. Identifying relevant studies.
3. Selecting studies.
4. Charting the data.
5. Collating, summarising, and reporting results.

A protocol was developed a priori to ensure methodological consistency, in line with best practice guidance for scoping review protocols [23]. It outlined the objectives, search strategy, and inclusion criteria and was followed throughout the process, although it was not registered externally.

In line with PRISMA-ScR guidance, the review did not include a formal assessment of study quality, as the aim was to map the existing evidence rather than evaluate the strength of individual studies.

3.2 Information Sources and Search Strategy

A comprehensive search was conducted across six academic databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SpringerLink, and ProQuest. Searches were conducted iteratively throughout April 2025 and finalised on 25 April 2025, covering studies published from January 2019 to April 2025. These databases were selected to ensure broad coverage of interdisciplinary research across cybersecurity, information systems, and human factors. Discipline-specific psychological databases such as PsycINFO and APA PsycExtra were not included. This decision was based on the primary anchoring of the review question in cybersecurity rather than psychology and on the expectation that the six selected databases index the most widely cited behavioural cybersecurity venues, including those that draw on psychological theories such as TPB, PMT, and TAM. However, we acknowledge that trust, risk perception, self-efficacy, and related

constructs are fundamentally psychological in nature, and that PsycINFO and APA PsycExtra may index relevant studies not captured by the selected databases. This omission is therefore a meaningful limitation for the human-focused part of the review, and replication studies should extend the search to these sources.

Search strings combined keywords related to trust and cybersecurity behaviour and were adapted to the syntax and indexing of each database. The general structure of the search query was as follows:

("trust" OR "trustworthiness") AND ("cybersecurity" OR "information security") AND ("risky behaviour" OR "security behaviour*" OR "user behaviour*" OR "compliance" OR "non-compliance" OR "protective behaviour*" OR "phishing" OR "password" OR "credential*" OR "security awareness" OR "adoption" OR "information security behaviour*")*.

A pilot run with the original three-cluster string (trust, cybersecurity, behaviour) and the extended string above was used to confirm that the broader behavioural synonyms did not introduce a substantively different set of eligible records, supporting the consistency of the search. While the core search terms were consistent across databases, minor variations were applied where necessary to account for differences in database requirements and terminology. To enhance coverage, the reference lists of included studies were screened manually using backward and forward citation searching to identify additional relevant papers. All retrieved records were exported to Microsoft Excel for management, deduplication, and screening.

3.3 Eligibility Criteria

Studies were included if they:

1. Examined trust (system, institutional, interpersonal, or generalised) in relation to cybersecurity behaviour.
2. Focused on adult populations such as employees, students, or general users where human participants were involved, or examined system-level, machine-learning, or testbed-based trust in relation to cybersecurity risk, system response, or security-related decision-making.
3. Reported empirical (quantitative, qualitative, or mixed methods) or theoretical research.
4. Were peer-reviewed journal articles, full conference papers, or doctoral dissertations/theses published in English between 2019 and 2025.

Studies were excluded if they:

1. Focused exclusively on technical trust mechanisms (e.g., cryptographic protocols) without a behavioural component.
2. Were commentaries, editorials, posters, or non-research items.
3. Did not discuss trust or cybersecurity behaviours.
4. Provided insufficient methodological detail.
5. Were duplicates.

During the screening process, inclusion and exclusion decisions were recorded systematically to support transparency and consistency. The full set of eligibility criteria is presented in Table 2.

Table 2: Inclusion and exclusion criteria for study selection.

Category	Inclusion Criteria	Exclusion Criteria/Rationale
Topic relevance	Studies that explicitly discuss trust (system, institutional, interpersonal, or generalised) in relation to cybersecurity behaviours, cybersecurity risk, system response, or security-related decision-making	Papers not addressing trust or not focused on cybersecurity behaviour, cybersecurity risk, or security-related decision-making
Population	Studies involving adult participants, such as employees, university students, professionals, or general users	Studies focused exclusively on children or adolescents
Study type	Empirical research, conceptual papers, theoretical papers, or system/testbed-based studies exploring trust in cybersecurity contexts	Non-research items such as editorials, commentaries, posters, or opinion pieces
Publication type	Peer-reviewed journal articles, full conference papers, and doctoral dissertations/theses published in English between 2019 and 2025. Doctoral dissertations/theses were included only where they provided sufficient methodological detail and directly addressed trust in relation to cybersecurity behaviour, cybersecurity risk, system response, or security-related decision-making	Non-English sources, studies published before 2019, non-academic sources, editorials, commentaries, posters, opinion pieces, and records lacking sufficient methodological detail
Focus of trust	Studies examining human-centred, socio-technical, institutional, interpersonal, generalised, or system-level trust where trust was linked to cybersecurity behaviour, cybersecurity risk, system response, or security-related decision-making	Studies focused solely on cryptographic, protocol-level, or mathematical trust mechanisms without behavioural, organisational, or cybersecurity risk-related relevance
Duplicates	If a study appeared more than once across databases, only the unique version was retained	Duplicate records were removed

3.4 Selection of Sources of Evidence

The database searches returned 3170 records. After removing five duplicates, 3165 titles and abstracts were screened, and 2990 records were excluded based on the predefined inclusion and exclusion criteria. A total of 175 articles were retrieved for full-text assessment, of which 130 were excluded. In line with PRISMA-ScR guidance [22], each excluded record was assigned a single primary reason for exclusion where more than one criterion applied. In order of frequency, the reasons were: the study did not examine trust or trustworthiness in a relevant sense; the study did not focus on cybersecurity behaviour, cybersecurity risk, or security-related decision-making; the population or study type did not meet the inclusion criteria; and the study lacked sufficient methodological detail to support data extraction. Where a study met more than one exclusion criterion, the primary reason was recorded. In total, 45 studies met all inclusion criteria and were included in the final analysis. Screening was conducted by the first author, with all uncertain cases referred to the supervisory team for a joint decision. The selection process is illustrated in the PRISMA flow diagram (Fig. 1).

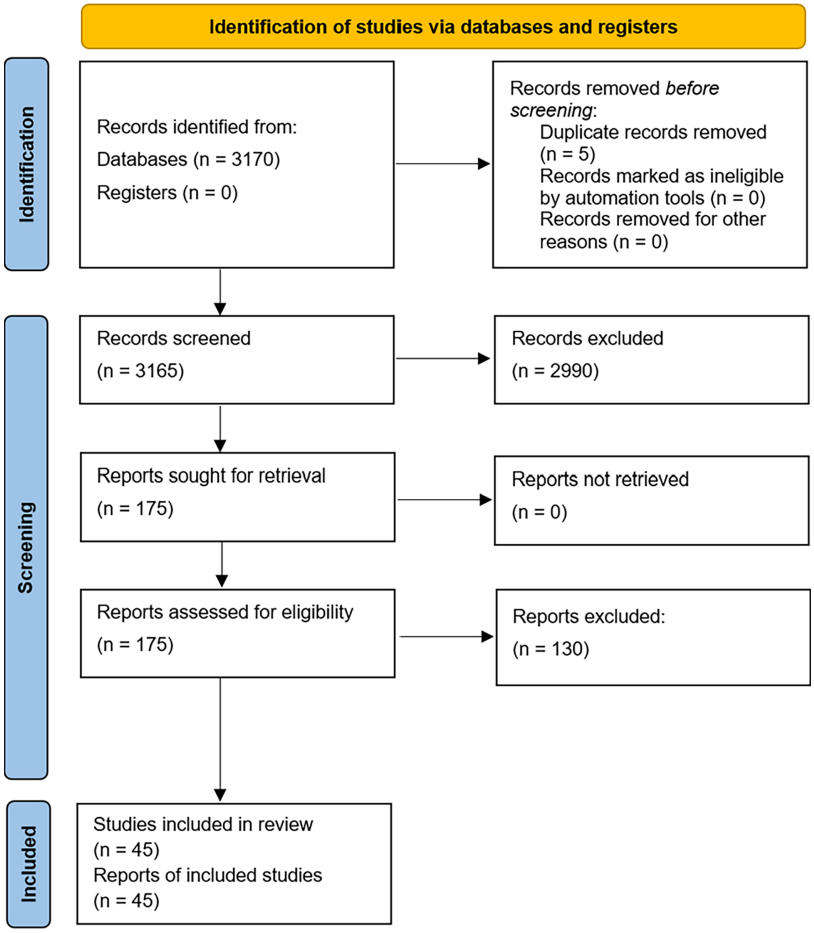

Figure 1: PRISMA-ScR flow diagram.

3.5 Data Charting and Extraction

Data from the 45 included studies were extracted using a structured charting form (Table 3). The charting form was developed iteratively and pilot tested on five papers to ensure clarity and consistency in data extraction. During piloting, ambiguities in variable definitions were identified and resolved before full extraction began. Extraction was conducted by the first author, and all uncertain or borderline cases were referred to the supervisory team for discussion and joint decision. This process was applied consistently across all 45 studies to support reliability and transparency in the charting process.

Table 3: Data extraction framework and variable definitions.

Category	Data Items Extracted	Description/Purpose
Bibliographic details	Author(s), year, study title, publication source	To identify publication patterns, research domains, and citation details
Study classification	Human-centred study or system/ML/testbed-based study	To distinguish studies involving human participants from system-level or technical studies

(Continued)

Table 3 (continued)

Category	Data Items Extracted	Description/Purpose
Geographic and contextual information	Country, region, sector, or setting	To map the geographical distribution and application context of the included studies
Study characteristics	Sample size, participant type, population group, and study design	To classify studies by population type and methodological approach
Trust constructs	System, institutional, interpersonal, or generalised trust	To determine how trust was conceptualised and operationalised across the literature
Behavioural focus	Adoption–compliance, social engineering, authentication, technical risk, or other/unclear behaviour	To classify the main cybersecurity behaviour or risk focus examined in each study
Theoretical frameworks	Theories or models, such as TPB, TAM/UTAUT, XAI, Mayer et al.'s trust model, socio-technical approaches, or other/none	To identify whether studies were guided by established theoretical or conceptual frameworks
Measurement approach	Self-report, system metric, or mixed/other measurement	To assess how trust, behaviour, or cybersecurity risk was measured
Effect direction	Positive, negative, mixed, or unclear	To capture how each study described the relationship between trust and cybersecurity behaviour or risk
Key findings and implications	Summary of the main results and conclusions reported by the authors	To capture evidence of how trust influences or relates to cybersecurity behaviour, risk, or system response
Identified gaps or limitations	Reported limitations, research gaps, or future directions	To synthesise emerging challenges and areas requiring further investigation

The following information was recorded for each study:

- Bibliographic details, including author, year, title, and publication source
- Country or context
- Region
- Sector or setting
- Study type, including whether the study was human-centred or system/ML/testbed-based
- Population or sample characteristics
- Study design
- Trust dimension, including system, institutional, interpersonal, or generalised trust
- Behaviour category, including adoption–compliance, social engineering, authentication, technical risk, or other/unclear behaviours
- Theoretical framework or model
- Measurement type, including self-report, system metric, or mixed/other
- Effect direction, including positive, negative, mixed, or unclear
- Key findings and implications
- Reported research gaps or limitations

A detailed study characteristics matrix is provided in Table 4, summarising the principal categorical characteristics of all 45 studies. The supplementary material also reports the full bibliographic details for all included studies, supporting transparency and replication.

3.6 Data Analysis and Synthesis

The synthesis combined descriptive analysis and thematic synthesis. Quantitative data were summarised using frequencies and percentages to identify patterns in study type, region, trust dimension, behaviour category, study design, measurement type, and effect direction. Descriptive findings were calculated for the full sample of included studies (N = 45). Where findings were related specifically to participant characteristics or human-focused cybersecurity behaviour, the analysis was based on the human-only subset. This distinction was applied to avoid mixing human-centred studies with system/ML/testbed-based studies.

For studies reporting a relationship between trust and cybersecurity behaviour or risk, the direction of the effect was coded based on the authors' own findings and interpretations, using a pre-defined codebook with explicit decision rules for each category. A positive (risk-increasing) effect was coded when higher trust was associated with increased risky behaviour, reduced compliance, or greater risk exposure. A negative (risk-reducing) effect was coded when higher trust was associated with reduced risky behaviour, improved compliance, the adoption of protective practices, or lower risk exposure. The terms "risk-increasing" and "risk-reducing" are used alongside "positive" and "negative" throughout the manuscript to reduce ambiguity. A mixed effect was coded where the relationship varied across contexts, variables, or outcomes. An unclear effect was coded when the study discussed trust but did not clearly report the direction of its relationship with cybersecurity behaviour or risk. It should be noted that, given the diversity of study designs and trust definitions across the included studies, some degree of interpretive judgement was unavoidable in applying these codes. This is a recognised limitation of thematic synthesis in scoping reviews and is reflected in the relatively high proportion of studies coded as unclear for effect direction (64.4%, n = 29).

In line with PRISMA-ScR guidance [22], the review did not include a formal assessment of study quality, as the aim was to map the existing evidence rather than evaluate the strength of individual studies. Screening was conducted by the first author, with all uncertain cases referred to the supervisory team for a joint decision, ensuring consistency in inclusion and exclusion decisions throughout the process.

Qualitative data were analysed thematically following the six-phase process [24]. This approach was selected because it provides a clear, transparent, and flexible framework for identifying and interpreting patterns across diverse qualitative findings. It also supports both inductive and deductive coding, enabling the integration of studies with varied designs.

Overall, this combined approach enabled both descriptive mapping and analytical interpretation of the literature, supporting the identification of key themes, relationships, and conceptual gaps.

Table 4: Study characteristics matrix.

No.	Author [Ref.]	Study Type		Study Design				Trust Dimension				Behaviour Category					Measurement Type				Region
		Human	System/ML/ Testbed	Qualitative	Mixed/Other	Experiment	Cross-sectional	System/ Testbed	Institutional	System	Interpersonal	Generalised	Adoption–Compliance	Social Engineering	Technical Risk	Authenti-cation	Other/ Unclear	Self-Report	System Metric	Mixed/Other	
1	Jayakumar et al. [25]		✓					✓		✓					✓				✓		Asia
2	Desri and Utomo [26]	✓					✓		✓				✓					✓			Asia
3	Apau et al. [27]	✓					✓		✓				✓					✓			Europe
4	Desai et al. [28]		✓		✓					✓					✓					✓	Global/ Multiple
5	Fernández-Martínez et al. [29]		✓					✓		✓					✓					✓	Europe
6	Ayushi et al. [30]	✓			✓				✓				✓							✓	Asia
7	Cao et al. [31]		✓					✓		✓							✓			✓	Global/ Multiple
8	Heydari and Nyarko [32]		✓					✓		✓							✓			✓	Northern America
9	Kaul et al. [33]		✓					✓		✓					✓					✓	Asia
10	Kumar et al. [34]		✓					✓		✓							✓			✓	Global/ Multiple
11	Nodehi et al. [35]		✓		✓					✓							✓			✓	Europe
12	Tam et al. [36]	✓			✓					✓			✓							✓	Europe
13	Zolanvari et al. [37]		✓					✓		✓							✓		✓		Global/ Multiple
14	Ramya and Azad [38]		✓					✓		✓							✓		✓		Asia
15	Han and Cho [39]		✓					✓		✓							✓		✓		Asia
16	Munir et al. [40]		✓					✓		✓							✓		✓		Northern America
17	Khan and Sharma [41]		✓					✓		✓							✓		✓		Asia
18	Siemers et al. [42]		✓		✓					✓							✓			✓	Europe

(Continued)

Table 4 (continued)

No.	Author [Ref.]	Study Type		Study Design				Trust Dimension				Behaviour Category					Measurement Type				Region
		Human	System/ML/ Testbed	Qualitative	Mixed/Other	Experiment	Cross-sectional	System/ Testbed	Institutional	System	Interpersonal	Generalised	Adoption–Compliance	Social Engineering	Technical Risk	Authentication	Other/ Unclear	Self-Report	System Metric	Mixed/Other	
19	Dong et al. [43]		✓					✓		✓							✓			✓	Asia
20	Neises et al. [44]	✓		✓						✓							✓	✓			Global/ Multiple
21	Kadena et al. [45]		✓		✓					✓						✓				✓	Europe
22	Greaves et al. [46]	✓					✓		✓				✓					✓			Europe
23	Khan et al. [47]		✓		✓					✓			✓							✓	Middle East/North Africa
24	Deljoo et al. [48]		✓					✓		✓							✓		✓		Europe
25	Ghernaouti et al. [49]	✓			✓				✓				✓							✓	Europe
26	Cellier and Ghernaouti [50]	✓			✓				✓				✓							✓	Europe
27	Markakis et al. [51]		✓		✓					✓			✓							✓	Europe
28	Kharvi [52]		✓					✓		✓							✓			✓	Northern America
29	Alhasan [53]	✓					✓				✓			✓				✓			Northern America
30	Kim [54]		✓					✓	✓				✓						✓		Northern America
31	Osunji [55]	✓		✓					✓				✓							✓	Africa
32	Yeoh et al. [56]		✓	✓						✓							✓			✓	Oceania
33	Pakaja [57]	✓					✓					✓	✓					✓			Asia
34	Palbar Misas et al. [58]	✓			✓					✓			✓							✓	Europe
35	Bajwa et al. [59]	✓					✓		✓				✓					✓			Asia
36	Itodo and Ozer [60]		✓		✓					✓			✓							✓	Global/ Multiple

(Continued)

Table 4 (continued)

No.	Author [Ref.]	Study Type		Study Design				Trust Dimension				Behaviour Category					Measurement Type				Region
		Human	System/ML/ Testbed	Qualitative	Mixed/Other	Experiment	Cross-sectional	System/ Testbed	Institutional	System	Interpersonal	Generalised	Adoption–Compliance	Social Engineering	Technical Risk	Authenti-cation	Other/ Unclear	Self-Report	System Metric	Mixed/Other	
37	van der Schyff and Flower-day [61]	✓					✓		✓					✓				✓			Global/ Multiple
38	Frank et al. [62]	✓			✓							✓					✓			✓	Global/ Multiple
39	Collett [63]		✓	✓					✓				✓							✓	Europe
40	Koohang et al. [6]	✓					✓		✓				✓					✓			Northern America
41	Butavicius et al. [64]	✓					✓			✓				✓						✓	Oceania
42	Sumithra et al. [65]		✓		✓					✓							✓			✓	Asia
43	Owen et al. [66]	✓					✓				✓			✓				✓			Africa
44	Vyhmeister and Castane [67]		✓		✓				✓								✓			✓	Europe
45	Razali et al. [68]	✓			✓						✓						✓			✓	Asia

Note: 1. Studies are cited by first author surname and Vancouver reference number [n]. "et al." is used for publications with more than two authors. Full bibliographic details are provided in the reference list. 2. Global/Multiple refers to studies conducted across more than one country.

4 Results

4.1 Overview of Included Studies

The publication trend (Fig. 2) shows a clear upward trajectory, with most studies appearing from 2022 onward. The peak in 2024 (n = 12) reflects the rise of system-level and machine-learning work on trust in cybersecurity, including AI-driven and IoT-focused testbeds [28,30,32–35]. The earlier years (2019–2021) include a higher share of policy-compliance and human-factors studies [6,48–51,64]. The studies published in 2025 (n = 6) included both human-centred and system- or framework-based work [25–27,52,57,67].

Geographically, Europe (31.1%) and Asia (26.7%) accounted for over half of the included studies, with a further 17.8% conducted across multiple countries (Global/Multiple). Northern America contributed 13.3%, while Africa (4.4%), Oceania (4.4%), and the Middle East/North Africa (2.2%) each had limited representation. This distribution shows that evidence on trust and cybersecurity behaviour remains concentrated in a small number of regions, with limited coverage from Africa, the Middle East, Latin America, and Oceania. Across the full dataset, 19 studies (42.2%) examined human or organisational behaviour, while 26 studies (57.8%) focused on system-level, machine-learning, or testbed-based contexts.

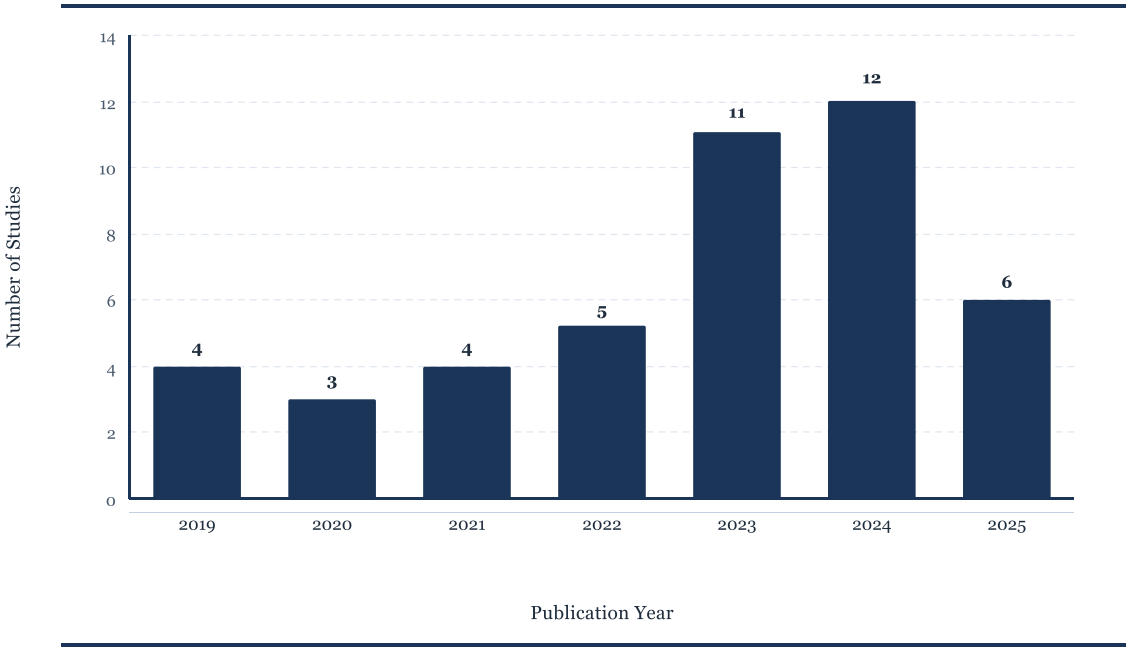

Figure 2: Publication trend (2019–2025).

The literature therefore covers both human-centred and system-oriented perspectives, but the system-level work currently forms the larger share. Where the analysis below depends on participant characteristics or human-focused behaviour, results are reported for the human-only subset (N = 19); descriptive characteristics are otherwise reported for the full sample (N = 45). Key characteristics of all included studies are summarised in the study characteristics matrix in Table 4, and a summary of methodological and contextual characteristics across all included studies is provided in Table 5.

Table 5: Summary of methodological and contextual characteristics of the 45 included studies (2019–2025).

Characteristic	Summary Distribution
Publication year	2019: 4 (8.9%); 2020: 3 (6.7%); 2021: 4 (8.9%); 2022: 5 (11.1%); 2023: 11 (24.4%); 2024: 12 (26.7%); 2025: 6 (13.3%)
Study classification	Human: 19 (42.2%); System/ML/Testbed: 26 (57.8%)
Geographic region	Europe: 14 (31.1%); Asia: 12 (26.7%); Global/Multiple: 8 (17.8%); Northern America: 6 (13.3%); Africa: 2 (4.4%); Middle East/North Africa: 1 (2.2%); Oceania: 2 (4.4%)
Study design	Mixed/Other: 16 (35.6%); System/Testbed: 15 (33.3%); Cross-Sectional: 10 (22.2%); Qualitative: 4 (8.9%)
Trust dimension	System: 27 (60.0%); Institutional: 13 (28.9%); Interpersonal: 3 (6.7%); Generalised: 2 (4.4%)
Behaviour category	Other/Unclear: 19 (42.2%); Adoption–Compliance: 17 (37.8%); Technical Risk: 4 (8.9%); Social Engineering: 4 (8.9%); Authentication: 1 (2.2%)
Theory/model	Other/None: 30 (66.7%); XAI: 5 (11.1%); TAM/UTAUT: 4 (8.9%); Mayer et al.: 2 (4.4%); Socio-Technical: 2 (4.4%); TPB: 2 (4.4%)
Measurement type	Mixed/Other: 27 (60.0%); Self-Report: 10 (22.2%); System Metric: 8 (17.8%)
Effect direction	Unclear: 29 (64.4%); Positive: 7 (15.6%); Negative: 5 (11.1%); Mixed: 4 (8.9%)

Focusing on the human-only studies (N = 19), participant groups were almost evenly split between general adult users (31.6%, n = 6) and professionals (31.6%, n = 6), followed by mixed or diverse cohorts (21.1%, n = 4). A further 15.8% (n = 3) did not specify the participant group. The human-focused literature, therefore, covers both general users and professional samples but offers limited representation of specific cohorts such as students or older adults, and a small share of studies still report participants in vague terms (see Fig. 3).

Cross-sectional surveys formed the largest proportion of the human-only sample (52.6%, n = 10), followed by mixed or other designs (36.8%, n = 7) and qualitative studies (10.5%, n = 2). No human-only study used an experimental or longitudinal design (see Fig. 4). This pattern is significant for several reasons. First, cross-sectional designs capture perceptions and reported behaviours at a single point in time, which means they cannot establish whether trust causes changes in cybersecurity behaviour, whether behaviour shapes trust, or whether both are driven by underlying factors such as security awareness, organisational culture, or prior experience with security incidents. Second, the dominance of self-report measurement in 52.6% of human-only studies introduces the possibility of social desirability bias, where participants report more secure behaviour than they actually exhibit. Third, the complete absence of experimental or longitudinal designs means the field currently lacks the evidence needed to test causal claims or to understand how trust in cybersecurity contexts develops, erodes, or can be shifted through targeted interventions. Together, these methodological characteristics limit what can be concluded from the available evidence and highlight the need for more diverse and robust research designs in future work.

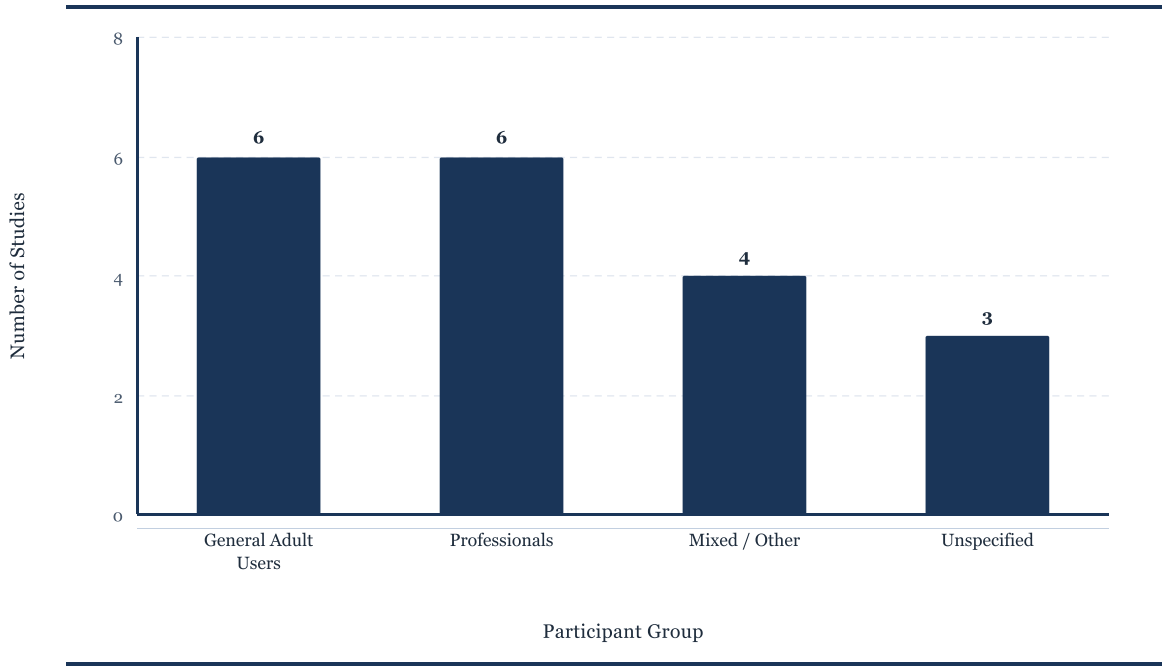

Figure 3: Population distribution (human-only, N = 19).

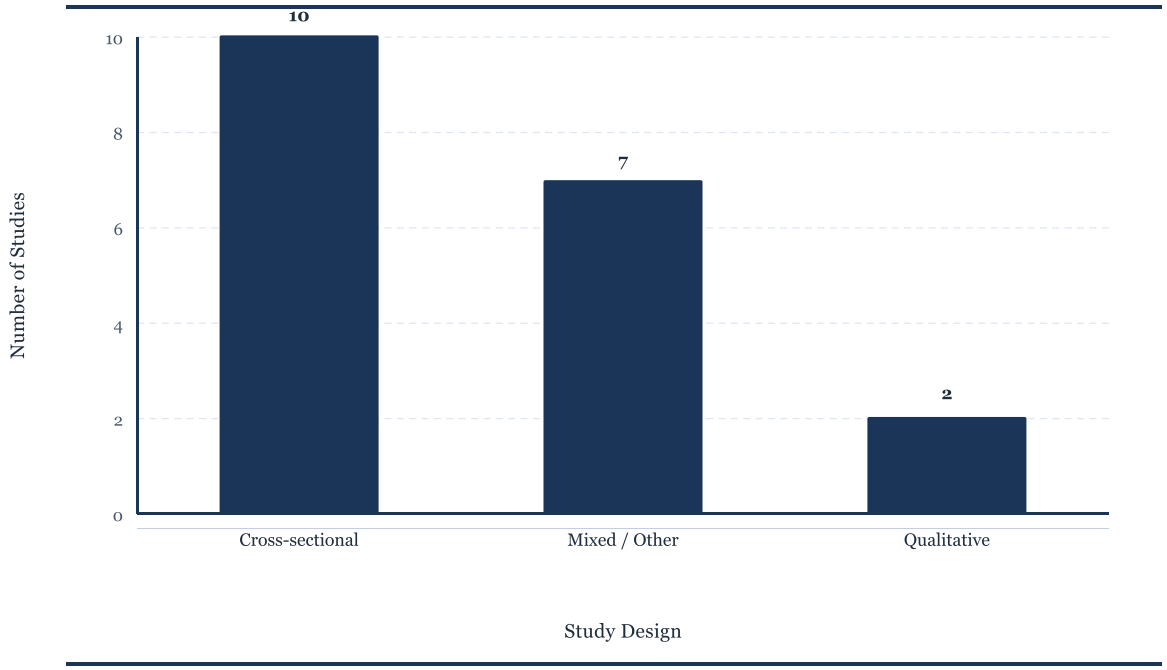

Figure 4: Study design distribution (human-only, N = 19).

4.2 Theme 1: Dimensions and Definitions of Trust

Trust was described in several ways, reflecting its multi-layered nature in cybersecurity research. This conceptual variation is not incidental but reflects the inherently context-dependent nature of trust as a construct. Studies in this review drawing on Mayer et al.'s trust model conceptualised trust through ability, benevolence, and integrity, with the relevance of these dimensions varying according to the trust relationship and context [6,51]. In cybersecurity settings, trust in a technical system draws primarily on perceptions of reliability and functionality, while trust in an organisation or institution involves assessments of competence and integrity, and trust in another person depends on interpersonal familiarity and perceived shared values. These distinctions mean that the definition and measurement of trust will necessarily vary across studies depending on the type of trust being examined, the population involved, and the cybersecurity context under investigation. The variation observed in the reviewed literature therefore reflects the multi-dimensional nature of trust rather than a lack of conceptual rigour [8,9].

- System trust (60.0%) refers to confidence in the reliability, security, or functionality of systems, algorithms, or platforms.
- Institutional trust (28.9%) relates to belief in the competence and integrity of organisations, service providers, or authorities.
- Interpersonal trust (6.7%) involves trust in other individuals, such as colleagues, peers, or online users.
- A small group of studies (4.4%) treated trust as a generalised construct without specifying a narrower trust dimension.

Overall, system trust was the most frequently examined dimension, followed by institutional trust, while interpersonal and generalised trust received much less attention (see Fig. 5). This pattern is largely driven by the system/ML/testbed studies, in which trust is operationalised as a numerical score or design property rather than a human perception [25,33,34,37,38]. When the analysis is restricted to the human-only subset (N = 19), institutional trust becomes the most frequently studied dimension (n = 10), followed by system trust (n = 4), interpersonal trust (n = 3), and generalised trust (n = 2). In other words, system trust dominates the literature mainly because system-level papers dominate the literature; in human-focused work the central construct is closer to institutional trust [6,55,58]. Across both subsets, trust is typically measured as a static property rather than a process that develops or erodes over time, which limits comparability across studies and makes it difficult to establish clear theoretical boundaries.

4.3 Theme 2: Cybersecurity Behaviours Associated with Trust

The reviewed studies examined a range of cybersecurity behaviours and security-related outcomes. These were grouped into five broad categories:

1. Adoption–compliance behaviours, such as following organisational policies, adopting security tools, or complying with security guidance.
2. Social engineering behaviours, including phishing, scams, and responses to trust cues used by attackers.
3. Authentication behaviours, such as password use, credential management, or two-factor authentication.
4. Technical risk behaviours, including malware-related activity, ignoring security warnings, unverified downloads, or weak technical security practices.
5. Other or unclear behaviours, where the behavioural focus was not clearly classifiable from the study information.

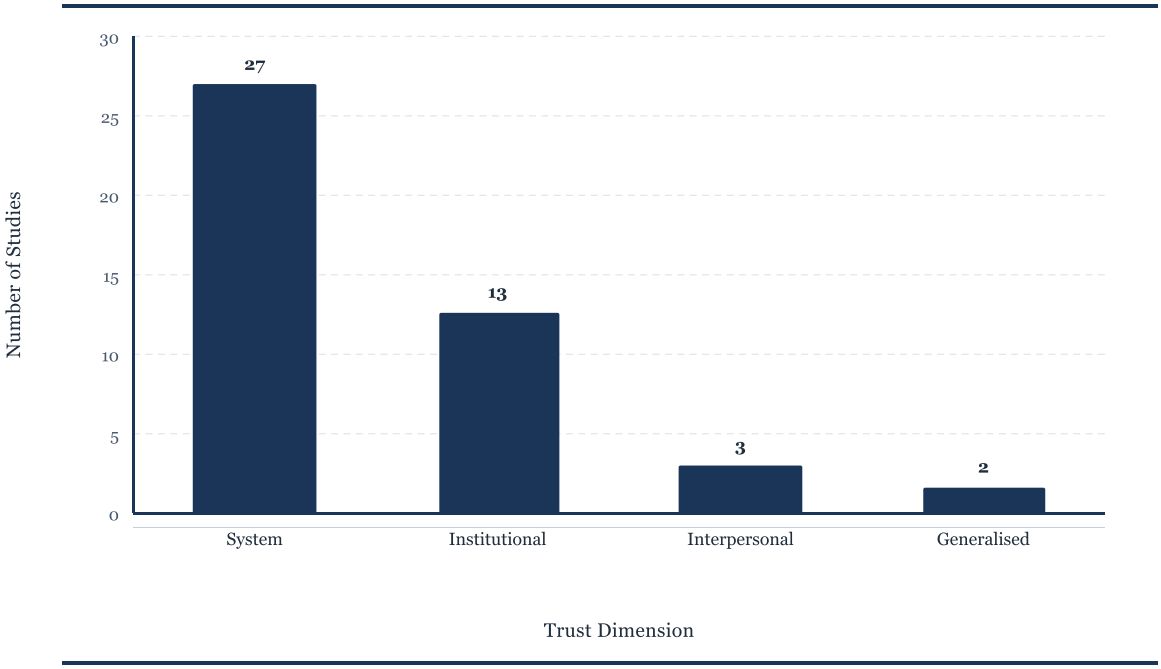

Figure 5: Trust dimensions (N = 45).

Across the full sample (N = 45), the most common behaviour category was other/unclear (42.2%, n = 19), followed by adoption-compliance (37.8%, n = 17). Technical risk and social engineering each accounted for 8.9% (n = 4), and authentication for 2.2% (n = 1). The high share of "other/unclear" is driven by system-level studies where the trust construct was operationalised at the device, model, or architectural level and the original authors did not link it to a specific human behaviour [25,29,37–41,43].

When the analysis is restricted to human-only studies (N = 19), adoption–compliance dominates (63.2%, n = 12), followed by social engineering (21.1%, n = 4) and other/unclear behaviours (15.8%, n = 3). Authentication and technical risk do not appear as the primary focus of any human-only study (see Fig. 6 and Table 4). The human-focused literature therefore concentrates on policy compliance, technology acceptance, and phishing-style decisions, while behaviours such as password practice, warning dismissal, or unverified downloads are addressed only indirectly or at the system level. This narrow behavioural footprint is itself a finding: it shows that the evidence supporting claims about "trust and cybersecurity behaviour" rests on a relatively small slice of behaviours, mainly compliance-related.

Overall, the findings suggest that trust is linked to different forms of cybersecurity behaviour. However, the type of behaviour examined varies across the literature, and several studies did not clearly define a specific behavioural category. This limits comparison across studies and highlights the need for clearer behavioural classification in future research.

4.4 Theme 3: Theoretical and Conceptual Foundations

In terms of theoretical grounding, 33.3% of the included studies applied a clear theoretical or conceptual framework, while 66.7% were coded as 'other' or 'none'. Among the studies that used an identifiable theory or model, the most common were XAI-related approaches (11.1%) and TAM/UTAUT (8.9%). Smaller numbers used Mayer et al.'s trust model (4.4%), socio-technical approaches (4.4%), or the Theory of Planned Behaviour (TPB) (4.4%).

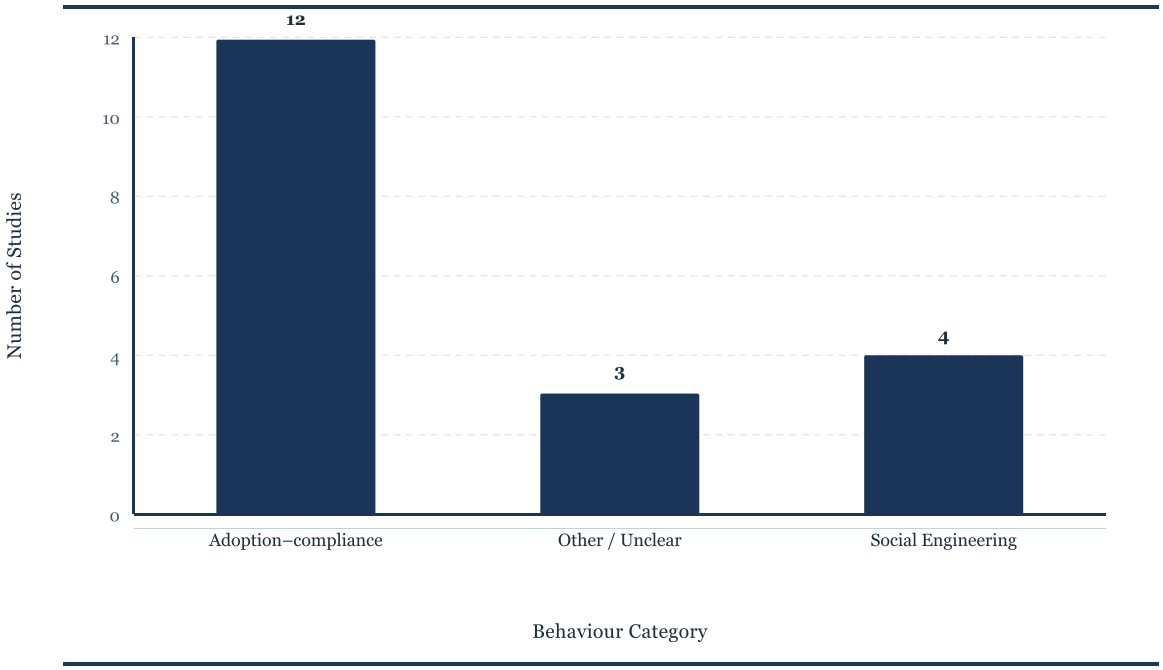

Figure 6: Behaviour categories (human-only, N = 19).

The theories that emerged addressed very different aspects of cybersecurity behaviour and trust, and they were rarely combined. XAI-related approaches and TAM/UTAUT focus on system use, transparency, automation, and user acceptance [28,32,37,40,60]; Mayer et al.'s trust model focuses on trust as a construct in its own right [6,51]; socio-technical approaches connect individual behaviour to organisational and technical context [44,49]; and the TPB [64] anchors the work in classical behavioural prediction. It is worth noting that XAI is more accurately a design paradigm than a behaviour theory, so the share of studies grounded in an explicit psychological or socio-technical framework is, in practice, closer to 20% than to 33.3%. The remaining 66.7% of studies (n = 30) did not apply a formal framework. Many of these reported descriptive or correlational findings or treated trust as an implicit construct embedded in a system metric. This limited theoretical grounding makes it difficult to explain how and why trust shapes cybersecurity behaviour and risk or to identify the conditions under which trust supports secure behaviour or contributes to vulnerabilities [69]. Stronger and more integrative use of behavioural, trust-based, and socio-technical theories would help close this gap and support the development of more predictive models in future research.

4.5 Theme 4: Methodological Characteristics and Effect Direction

The methodological characteristics of the reviewed studies varied widely. Across all 45 studies, the most common study designs were mixed/other designs (35.6%) and system/testbed-based studies (33.3%). Cross-sectional studies made up 22.2%, qualitative studies made up 8.9%, and no experimental studies were identified. This distribution reflects the mixed nature of the literature, which includes both human-focused empirical studies and system-level or testbed-based research.

Across the full sample, the most common measurement type was mixed/other (60.0%), followed by self-report (22.2%) and system metrics (17.8%). In the human-only subset (N = 19), self-report was used in 52.6% of studies, while 47.4% were coded as mixed/other. This shows that self-report remains important in the human-focused literature, although it is not the dominant measurement type across the full sample.

Focusing specifically on the human-only studies (N = 19), the direction of trust–behaviour relationships was uneven (Fig. 7).

The most common category was unclear (31.6%, n = 6), followed by negative (26.3%, n = 5), while positive and mixed effects each accounted for 21.1% (n = 4). In this review, the codebook defines a positive effect as higher trust being associated with greater risky behaviour or reduced compliance, while a negative effect indicates that higher trust is linked to safer behaviour or stronger compliance. This polarity differs from intuitive readings of "positive/negative", so it should be interpreted in line with the codebook rather than at face value.

Across all 45 studies the picture is even less conclusive: 29 studies (64.4%) reported unclear effects, 7 studies (15.6%) reported positive effects, 4 studies (8.9%) reported mixed effects, and 5 studies (11.1%) reported negative effects. The dominance of "unclear" reflects the inclusion of system/ML/testbed papers [25,29,37–39,41,43], where the direction of the effect on human behaviour is not defined. Once these are removed, the human-only evidence shows that clearly directional findings were reported in 9 of the 19 studies, while the remaining 10 studies reported mixed or unclear effects. The relationship between trust and cybersecurity behaviour is therefore inconsistent and context-dependent based on the available evidence, and many studies do not report a clear directional effect.

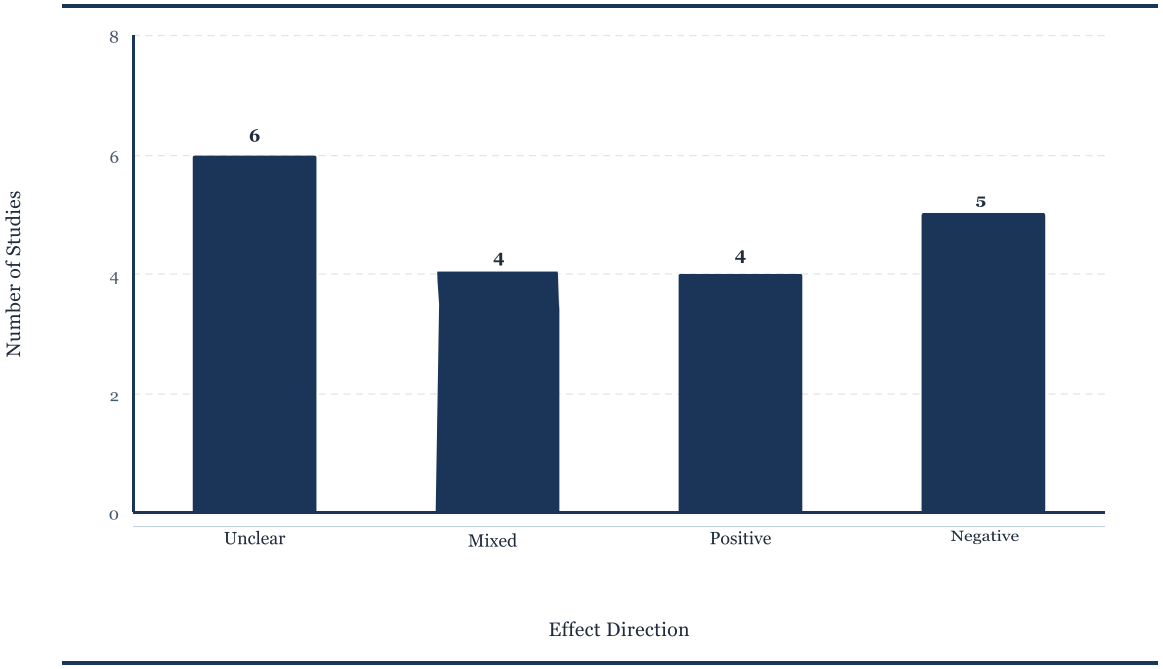

Figure 7: Effect direction (human-only, N = 19).

4.6 Theme 5: Influencing Factors and Research Gaps

Several factors appeared to influence the relationship between trust and cybersecurity behaviour. These include perceived risk, security awareness, self-efficacy, organisational culture, technology anxiety, and previous experience with security incidents. However, only a small number of studies examined these factors directly, and most did so in a limited or descriptive manner.

Overall, these gaps highlight the need for more theory-driven, methodologically balanced, and context-aware research to develop a clearer and more coherent understanding of how trust shapes cybersecurity behaviour. These gaps are summarised in Table 6.

Table 6: Summary of identified research gaps.

Gap No.	Research Gap	Description/Evidence from Review
G1	Limited theoretical integration	66.7% of included studies (n = 30) did not apply a clear or unified theoretical framework linking trust, risk, and cybersecurity behaviour. Among those that did, theories were rarely combined, with XAI-related approaches (11.1%), TAM/UTAUT (8.9%), and TPB (4.4%) applied separately rather than in an integrated manner [6,36,64].
G2	Methodological imbalance	In the human-only subset (N = 19), 52.6% of studies used cross-sectional designs, 36.8% used mixed or other designs, and 10.5% used qualitative designs, with no experimental or longitudinal studies identified. In terms of measurement, self-report was used in 52.6% of the human-only studies [6,53,66].
G3	Limited focus on interpersonal and contextual trust	Across the full sample (N = 45), system trust accounted for 60.0% of studies and institutional trust for 28.9%, while interpersonal trust was examined in only 6.7% (n = 3) and generalised trust in 4.4% (n = 2). Peer influence, group dynamics, and situational factors were infrequently addressed [53,61,62].
G4	Weak operationalisation of trust	Studies varied considerably in how trust was defined and measured, using different dimensions, scales, and indicators across human-centred and system-level contexts. This variability is reflected in the high proportion of studies coded as unclear for effect direction (64.4%, n = 29), making direct comparison across studies difficult [36,64,68].
G5	Limited integration between system-level and human-centred evidence	System/ML/testbed studies accounted for 57.8% of the sample (n = 26), while human-centred studies accounted for 42.2% (n = 19). These two strands rarely intersected within the same study. Several system-level studies proposed trust scoring or zero-trust architectures without measuring how these signals shaped users' subsequent cybersecurity decisions [37,56,60].
G6	Lack of longitudinal and intervention research	No human-only study in this review used an experimental or longitudinal design. Very few studies examined how trust develops or changes over time, or assessed the causal effects of training, system design, or policy interventions on cybersecurity behaviour. This represents a critical gap for building an evidence base capable of supporting practical recommendations [6,59,64].

4.7 Summary of Findings

Overall, the review shows that research on trust in cybersecurity is growing but uneven. System and institutional trust receive thorough coverage, especially in technical and policy-compliance work, while interpersonal and generalised trust remain underexplored. The literature includes both human-centred and system-level evidence, but these strands are rarely connected within the same study, and the direction

of the trust–behaviour relationship is unclear or mixed in roughly three-quarters of the included papers. Trust therefore appears as a context-dependent factor that can support or weaken cybersecurity, rather than a consistently positive or negative influence. These patterns motivate the more integrated reading of the evidence developed in the following discussion, which links the themes above to the proposed conceptual framework.

5 Discussion

This section summarises the key patterns identified in the review and interprets them in relation to the proposed conceptual framework. Fig. 8 presents an integrated view of the relationships among trust dimensions, cybersecurity behaviours, and influencing factors identified across the reviewed studies. The framework is used to guide the interpretation of the findings and to show how trust may shape cybersecurity behaviour in both human-centred and system-level contexts.

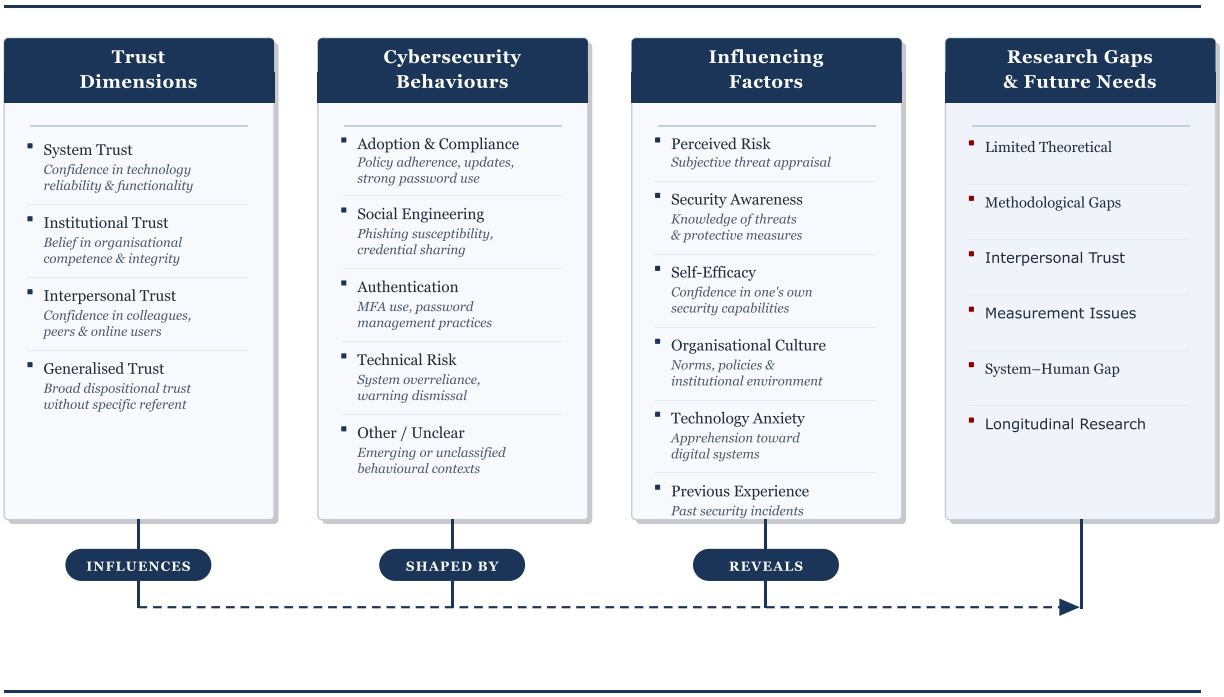

Figure 8: Conceptual framework synthesising the relationship between trust and cybersecurity behaviour.

The framework illustrates how system, institutional, interpersonal, and generalised trust relate to different cybersecurity behaviours and security-related outcomes. These relationships may also be shaped by factors such as awareness, perceived risk, self-efficacy, organisational culture, and technology-related experience. Rather than presenting trust as consistently positive or negative, the framework shows that the evidence is mixed and context-dependent. To move beyond a descriptive mapping, the framework is presented here as a set of working propositions that future studies can test, refine, or refute. Three propositions follow most directly from the reviewed evidence.

P1. Institutional trust is most clearly linked to adoption and compliance behaviours in the reviewed human-focused literature. Studies that examine organisational governance, leadership, and policy clarity report institutional trust as a meaningful factor in compliant decision-making [6,51].

P2. System trust, when poorly calibrated, can produce over-reliance on technical controls, which several studies link to riskier cybersecurity decisions [64]. A related calibration issue, in the form of overconfidence in security knowledge or judgement, is reflected in studies that examine generalised trust and behavioural bias [57,62].

P3. Interpersonal and generalised trust shape susceptibility to social engineering, with cues that imitate familiar senders or in-group identities reportedly increasing the willingness to take security risks [53,66].

Among the relationships represented in the framework, the links between institutional trust and adoption-compliance, and between system trust and risky decision-making in technical contexts, are the most clearly supported by the reviewed studies. The links involving interpersonal and generalised trust, and the connection between system-level trust signals and human cybersecurity behaviour, are plausible but currently underexamined in the literature; in this review they are presented as priorities for future empirical work rather than as established findings. The framework is therefore a structured and testable map of where the evidence is stronger and where future work is needed, rather than a closed model.

The following subsections extend this framework by discussing the thematic insights in more detail, including the dual role of trust, the dominance of system and institutional perspectives, theoretical limitations, methodological issues, emerging contexts, and practical implications.

5.1 The Dual Role of Trust

The findings indicate that trust can act as a double-edged factor in cybersecurity. When trust is well calibrated, it supports cooperation, encourages compliance, and increases users' confidence in security technologies. For example, appropriate trust in organisational security tools can help employees follow recommended practices, whereas misplaced trust, such as assuming that a familiar-looking email or automated alert is always accurate, may lead users to click on malicious links or overlook threats.

However, the empirical evidence does not support a single direction of influence. In the human-only subset, 4 studies (21.1%) reported a positive effect, 5 (26.3%) reported a negative effect, and 10 (52.6%) reported mixed or unclear effects. The role of trust therefore depends on the trust dimension being examined, the behaviour being studied, and the context in which trust is formed. Some studies linked trust to over-reliance and reduced compliance [62,70], while others suggested that institutional or interpersonal trust supports safer behaviour [6,51]; many simply could not establish a clear direction.

This tension reflects the broader balance between security and usability. Users who overtrust systems may overlook potential threats, while those with very low trust may ignore or avoid protective tools. These findings highlight trust calibration as an important direction for future cybersecurity research [70,71]. Supporting users in developing an appropriate level of trust, aligned with the actual capability of security tools, may help improve security decision-making and reduce risky behaviour without encouraging either unnecessary distrust or overconfidence.

5.2 The Dominance of System and Institutional Perspectives

System trust dominates the literature mainly because system-level studies also dominate it. As Table 4 shows, 26 of the 45 included papers (57.8%) are system, ML, or testbed studies, and the trust construct in these papers is typically a property of a model, device, or architecture rather than a property of a person. Once those studies are removed, institutional trust becomes the most frequently examined dimension in the human-only subset. The visible "system trust" pattern in the headline figures is therefore partly an artefact of the methodological mix, and not only a substantive finding.

By contrast, interpersonal and generalised trust have received much less attention. This is important because systems and institutions do not shape cybersecurity decisions alone. Social relationships, workplace norms, peer advice, and broader assumptions about whether people or systems can be trusted may also influence them.

For example, employees may rely on colleagues when making quick security-related decisions, such as judging whether an unexpected email is safe or whether it is acceptable to share a file or password [72]. Similarly, generalised trust may influence how users respond to unfamiliar systems, online services, or digital requests. These forms of trust are relevant to cybersecurity behaviour, but they remain less developed in the existing literature.

These findings suggest greater attention to interpersonal and generalised forms of trust. Incorporating these dimensions into future research would support a more comprehensive and human-centred understanding of cybersecurity behaviour, as reflected in the proposed conceptual framework.

5.3 Limited Theoretical Integration

The findings show that 66.7% of the reviewed studies (n = 30) lacked a clear theoretical or conceptual framework, and the 33.3% that did use theories that were applied separately rather than integrated. The most common identifiable approaches included XAI-related perspectives, TAM/UTAUT, Mayer et al.'s trust model, socio-technical approaches, and TPB.

This limited theoretical integration makes it difficult to explain how and why trust influences cybersecurity behaviour and cybersecurity-related risk. For example, technology acceptance models can help explain why users adopt or rely on a system, while trust models can explain confidence, vulnerability, and perceived reliability. Socio-technical approaches can also help connect individual behaviour with organisational and technical contexts. However, researchers rarely bring these perspectives together in a single analytical framework.

As a result, many studies describe trust-related outcomes without fully explaining the mechanisms behind them. This is particularly important because the findings show that trust does not have one consistent effect. It may support secure behaviour in some contexts, but it may also increase the risk when it leads to over-reliance, reduced caution, or misplaced confidence.

Greater integration of behavioural, trust-based, and socio-technical theories would provide a stronger foundation for future research. It would also help explain when trust supports cybersecurity and when it becomes a source of vulnerability.

5.4 Methodological Challenges

The methodological profile of the reviewed studies was varied, reflecting the broad scope of research on trust and cybersecurity. The full sample included system/testbed-based studies, mixed or other designs, cross-sectional studies, and qualitative studies. This variation shows that trust has been examined across both human-centred and system-level contexts.

However, the human-focused evidence remains methodologically thin. In the human-only subset (N = 19), cross-sectional designs accounted for 52.6% of studies (n = 10), followed by mixed or other designs (36.8%, n = 7) and qualitative designs (10.5%, n = 2). No human-only study used an experimental or longitudinal design. In terms of measurement type, self-report was used in 52.6% of human-only studies (n = 10), while 47.4% (n = 9) relied on mixed or other measures.

These methodological characteristics limit what can be concluded from the existing evidence. Cross-sectional and self-report-based approaches are useful for exploring perceptions, attitudes, and reported

behaviours, but they cannot show whether trust leads to behaviour, whether behaviour shapes trust, or whether both are driven by other factors such as awareness, perceived risk, or organisational culture.

Future studies would benefit from more diverse and robust methodologies. Longitudinal designs could help explain how trust changes over time, while experimental and intervention-based studies could provide stronger evidence about how trust influences behaviour in specific cybersecurity contexts. Behavioural observation, system log analysis, and mixed-method designs could also support a clearer understanding of how trust operates in real-world security decisions [73].

Overall, the methodological imbalance identified in this review calls for stronger research designs that can connect trust, behaviour, and context more clearly.

5.5 Emerging Directions and New Contexts

The reviewed literature shows growing attention to trust in emerging technological contexts, including AI, automation, explainable AI, the Internet of Things (IoT), and system-level cybersecurity models. These studies are important because they show how researchers are increasingly examining trust beyond traditional user behaviour and organisational compliance.

However, the connection between system-level trust and human cybersecurity behaviour remains underdeveloped. Several studies examine trust in technical systems, machine learning models, or testbed environments, but fewer studies clearly explain how these forms of trust affect users' actual cybersecurity decisions. This creates a gap between technical understandings of trust and human-centred behavioural evidence. More specifically, several studies in this review proposed system-level trust scoring or zero-trust architectures [25,31,33,34,37], but none of them measured how those signals shaped users' subsequent cybersecurity decisions, illustrating the system–human gap noted in Theme 5.

For example, users may rely on AI-generated alerts or recommendations without fully verifying their accuracy, reflecting automation bias and over-reliance on AI outputs [74,75]. In other cases, users may avoid or ignore security technologies if they lack an understanding of how these systems work or do not trust their outputs. These examples show that trust in emerging technologies can shape behaviour in different ways, depending on system design, transparency, user awareness, and context.

Further research is needed to better connect technical and human-centred perspectives. This includes examining how users form, adjust, and calibrate trust in automated systems and how factors such as transparency, explanation, feedback, and previous experience influence cybersecurity behaviour. This need for stronger system-human integration is also reflected in the proposed conceptual framework (see Fig. 8).

5.6 Practical Implications

The findings have several practical implications for organisations, system designers, and cybersecurity educators.

First, cybersecurity awareness and training programmes should include guidance on trust calibration [70]. Users should not simply be encouraged to trust or distrust systems, but should be supported in judging when trust is appropriate and when caution is needed. For example, training can show how phishing messages imitate trusted sources [64], or how automated tools may still produce errors that require human judgement [61,62]. Butavicius et al.'s "trust in technical controls" scale [64] suggests that compliance training should differentiate between trust in tools and trust in organisations, rather than treating them as a single construct.

Second, organisations can strengthen trust in cybersecurity practices by promoting transparency, clear communication, senior management support, and consistent governance practices [76]. When users

understand how security policies are developed, why certain controls are required, and how their data is protected, they may be more likely to engage with cybersecurity practices in a meaningful way.

Third, the design of AI-enabled, automated, and system-level security tools should support informed trust. This means providing users with clear explanations, useful feedback, and enough information to evaluate system recommendations without encouraging over-reliance. In this sense, technical reliability should be combined with human-centred design.

Finally, the findings suggest that cybersecurity strategies should not treat trust as a single or fixed factor. System, institutional, interpersonal, and generalised trust may influence behaviour in different ways. Therefore, organisations should consider how trust operates across technical, organisational, and social contexts when designing policies, training, and security tools.

Together, these approaches connect technical reliability with psychological, organisational, and behavioural factors, supporting the development of more resilient and trustworthy human-technology environments.

5.7 Conceptual and Methodological Gaps

Bringing these threads together, the review identifies a coherent set of conceptual and methodological gaps. The literature lacks integrative frameworks that connect trust, risk, and behaviour; relies heavily on cross-sectional and self-report data in the human-focused subset; treats interpersonal and generalised trust as marginal; defines and measures trust inconsistently; and rarely links system-level trust signals to human cybersecurity decisions. The conceptual framework in Fig. 8 is intended as a structuring response to these gaps rather than a closed model: it positions system, institutional, interpersonal, and generalised trust as the input dimensions, the five behavioural categories as the outputs, and the influencing factors and research gaps as the contextual layer that future work needs to address with stronger theory, more diverse methods, and tighter system–human integration.

6 Limitations and Future Work

6.1 Open Research Issues

Beyond the methodological limitations discussed below, the review identifies four open research issues that represent unresolved theoretical and empirical problems in the current literature. These are distinct from future research directions in that they reflect fundamental questions the field has not yet answered, rather than methodological recommendations for how future studies should be designed.

OI1. The trust calibration problem. When is trust in a cybersecurity system, institution, or person appropriate, and when does it become a source of vulnerability? The reviewed evidence shows that trust can both support and undermine cybersecurity behaviour, depending on context, but no study clearly identifies the conditions under which trust transitions from protective to risk-increasing. Developing theoretically grounded criteria for appropriate trust calibration remains an open problem [70,71].

OI2. The measurement problem. What constitutes a valid, reliable, and comparable measure of trust across different cybersecurity contexts? The reviewed studies used varied dimensions, scales, and indicators, making direct comparison difficult and effect directions unclear in 64.4% of cases. A shared measurement framework for trust in cybersecurity research does not yet exist [69].

OI3. The interpersonal trust gap. How do peer relationships, social norms, and group dynamics shape cybersecurity decisions in ways that system and institutional trust cannot explain? Only 6.7% of studies (n = 3) examined interpersonal trust, and the mechanisms through which colleagues, peers, and social networks influence security behaviour remain poorly understood [72].

OI4. The system-human disconnect. How, if at all, do system-level trust signals translate into human cybersecurity behaviour? System/ML/testbed studies accounted for 57.8% of the sample, yet none of the reviewed studies clearly measured how technical trust signals, such as zero-trust architectures or AI-generated alerts, shaped users' subsequent security decisions [74,75].

6.2 Limitations

Although this review followed a structured and transparent process, several limitations should be acknowledged.

First, the review included only English-language academic sources published between 2019 and 2025, including peer-reviewed journal articles, full conference papers, and a small number of doctoral dissertations/theses. Although dissertations can provide detailed methodological and contextual evidence, their inclusion may introduce variation in publication type, and this should be considered when interpreting the findings.

Second, the search strategy focused on major academic databases, so studies indexed in smaller or more specialised sources may not have been captured. In particular, PsycINFO and APA PsycExtra were not searched; this omission is a meaningful limitation for the human-focused part of the review, and replication studies should extend the search to these sources.

Third, the analysis depended on the information reported by the original authors. Inconsistencies in the definitions or measurements of trust and cybersecurity behaviour, missing methodological details, and ambiguous descriptions could not always be resolved during data extraction. This variability is reflected in the relatively high proportion of studies coded as "unclear" for effect direction (64.4% across the full sample and 31.6% in the human-only subset), which limits the strength of any direct claim about whether trust supports or undermines cybersecurity behaviour.

Fourth, the human-only subset is small (N = 19) and is dominated by cross-sectional and self-report designs, with no experimental or longitudinal studies. Conclusions about human-focused patterns should therefore be treated as indicative rather than definitive. The included studies also varied considerably in design, context, and measurement approach: some focused on human participants, while others examined system-level, machine learning, or testbed-based trust. This breadth was useful for mapping the field but limited direct comparison across studies.

In line with PRISMA-ScR guidance, the review excluded a formal critical appraisal or quantitative meta-analysis; this limitation is a methodological boundary of the scoping approach rather than a flaw, but it does mean that study quality and effect sizes were not formally assessed. Future research should address these limitations by including non-English studies and grey literature where appropriate, applying systematic or meta-analytic approaches once the evidence base becomes more consistent, and adopting experimental or longitudinal designs to strengthen causal understanding [77]. Stronger attention is also needed to cross-cultural perspectives [78], interpersonal dynamics, and the calibration of trust in emerging technological contexts such as artificial intelligence, automation, and the Internet of Things [79]. Greater geographical diversity is also needed, with underrepresented regions such as Africa, the Middle East, and Latin America offering important and largely unexplored contextual variation in how trust shapes cybersecurity behaviour [80]. A specific priority is to better connect system-level and human-centred perspectives so that trust in technical systems, automated tools, and machine learning models can be explicitly linked to the cybersecurity decisions and behaviours that follow from them.

Together, these directions would support stronger theoretical integration, improve methodological rigour, and provide a more robust foundation for developing human-centred and trustworthy cybersecurity systems.

7 Conclusion

This scoping review mapped and summarised current evidence on how trust relates to cybersecurity behaviour and cybersecurity-related risk. Drawing on 45 studies published between 2019 and 2025, including a human-only subset of 19, the findings show that trust is a complex socio-technical factor that can both support and weaken cybersecurity, depending on how it is formed, maintained, and applied.

Three patterns stand out, although the human-focused subset is small (N = 19) and the conclusions about that subset should be read as indicative rather than definitive. First, the literature is heavily weighted towards system and institutional trust, while interpersonal and generalised trust remain underexplored, particularly in human-focused work. Second, the evidence covers both human-centred and system-level studies, but these strands rarely intersect within the same paper, limiting what can be said about how technical trust signals translate into user behaviour. Third, the direction of the trust–behaviour relationship is not consistent across studies: some studies linked trust to increased risk or reduced compliance, others suggested that trust supports safer behaviour, and a large share reported mixed or unclear effects. This last pattern signals a need for stronger theoretical integration and clearer, more comparable measurement approaches rather than a settled empirical claim about trust.

Overall, the evidence reinforces the view of trust as a dynamic and context-dependent factor in cybersecurity. Future research should apply stronger theoretical foundations, adopt more robust and diverse methodologies, and connect system-level trust signals more explicitly to human cybersecurity behaviour.

In practice, the review suggests that organisations and system designers should focus on calibrating trust rather than simply increasing it. Training and awareness programmes can help individuals recognise when to trust systems, organisations, or other people and when to exercise caution. Clear communication, explainable AI, transparent security policies, and human-centred system design can support informed and adaptive trust, helping to reduce vulnerability to human error.

Acknowledgement: Not applicable.

Funding Statement: This doctoral research is funded through a scholarship provided by the Ministry of Education, Saudi Arabia, and supported by the Saudi Arabian Cultural Bureau in London. No specific grant number was assigned to this sponsorship.

Author Contributions: Shadi Melebari was responsible for the conceptualisation, methodology, formal analysis, investigation, data curation, original draft writing, and visualisation. Muhammad Atif Ur Rehman, Ali Kashif Bashir, and Mohammed Al-Khalidi contributed to supervision, validation, and review and editing of the manuscript. Muhammad Atif Ur Rehman also supported project administration. Shadi Melebari was responsible for funding acquisition. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting this review are derived from the included studies and are summarised in the tables and supplementary materials. No primary empirical participant data were generated for this study.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at https://www.techscience.com/doi/10.32604/cmc.2026.085515/s1.

References

1. Rahman T, Rohan R, Pal D, Kanthamanon P. Human factors in cybersecurity: a scoping review. In: The 12th International Conference on Advances in Information Technology. Bangkok, Thailand: ACM; 2021. p. 1–11.
2. Hoff K, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors. 2015;57(3):407–34. doi:10.1177/0018720814547570.
3. Khando K, Gao S, Islam SM, Salman A. Enhancing employees information security awareness in private and public organisations: a systematic literature review. Comput Secur. 2021;106:102267. doi:10.1016/j.cose.2021.102267.
4. De Kimpe L, Walrave M, Verdegem P, Ponnet K. What we think we know about cybersecurity: an investigation of the relationship between perceived knowledge, internet trust, and protection motivation in a cybercrime context. Behav Inf Technol. 2022;41(8):1796–808.
5. Morice D. Trust framework on exploitation of humans as the weakest link in cybersecurity. Appl Cybersecur Internet Gov. 2023;2(1):1–26. doi:10.60097/acig/162867.
6. Koohang A, Nowak A, Paliszkiewicz J, Nord JH. Information security policy compliance: leadership, trust, role values, and awareness. J Comput Inf Syst. 2020;60(1):1–8. doi:10.1080/08874417.2019.1668738.
7. Khadka K, Ullah AB. Human factors in cybersecurity: an interdisciplinary review and framework proposal. Int J Inf Secur. 2025;24(3):119. doi:10.1007/s10207-025-01032-0.
8. Uttenthal M. A conceptual analysis of trust. Soc Sci Inf. 2024;63(3):392–410. doi:10.1177/05390184241270835.
9. Bodó B. Mediated trust: a theoretical framework to address the trustworthiness of technological trust mediators. New Media Soc. 2021;23(9):2668–90. doi:10.1177/1461444820939922.
10. Spadaro G, Gangl K, Prooijen JWV, Lange PAMV, Mosso CO. Enhancing feelings of security: how institutional trust promotes interpersonal trust. PLoS One. 2020;15(9):e0237934. doi:10.1371/journal.pone.0237934.
11. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991;50(2):179–211. doi:10.1016/0749-5978(91)90020-t.
12. Rogers RW, Cacioppo J, Petty R. Cognitive and physiological processes in fear appeals and attitude change: a revised theory of protection motivation. In: Social psychophysiology: a sourcebook. New York, NY, USA: Guilford Press; 1983. p. 153–77.
13. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q. 1989;13(3):319–40. doi:10.2307/249008.
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19–32. doi:10.1080/1364557032000119616.
15. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5:69. doi:10.1186/1748-5908-5-69.
16. Ali RF, Dominic PDD, Ali SEA, Rehman M, Sohail A. Information security behavior and information security policy compliance: a systematic literature review for identifying the transformation process from noncompliance to compliance. Appl Sci. 2021;11(8):3383. doi:10.3390/app11083383.
17. Almansoori A, Al-Emran M, Shaalan K. Exploring the frontiers of cybersecurity behavior: a systematic review of studies and theories. Appl Sci. 2023;13(9):5700.
18. Angraini, Alias RA, Okfalisa. Information security policy compliance: systematic literature review. Procedia Comput Sci. 2019;161:1216–24. doi:10.1016/j.procs.2019.11.235.
19. Nifakos S, Chandramouli K, Nikolaou CK, Papachristou P, Koch S, Panaousis E, et al. Influence of human factors on cyber security within healthcare organisations: a systematic review. Sensors. 2021;21(15):5119. doi:10.3390/s21155119.
20. Wang Z, Zhu H, Sun L. Social engineering in cybersecurity: effect mechanisms, human vulnerabilities and attack methods. IEEE Access. 2021;9:11895–910. doi:10.1109/access.2020.2992807.
21. Kannelønning K, Katsikas SK. A systematic literature review of how cybersecurity-related behavior has been assessed. Inf Comput Secur. 2023;31(4):463–77. doi:10.1108/ics-08-2022-0139.

22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73. doi:10.7326/m18-0850.
23. Peters MDJ, Godfrey C, McInerney P, Khalil H, Larsen P, Marnie C, et al. Best practice guidance and reporting items for the development of scoping review protocols. JBI Evid Synth. 2022;20(4):953. doi:10.11124/jbies-21-00242.
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. doi:10.1191/1478088706qp063oa.
25. Jayakumar D, Manickavalli N, Dhatchayini K, Ramkumar MO, Rajmohan R. Enhancing security and trust management in IoT-driven edge computing environment using ensemble learning techniques. In: Proceedings of the 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN); 2025 May 14–16; Salem, India. p. 986–91.
26. Desri RA, Utomo RG. Mobile banking security: relationship between trust, risk perception, and security based on UTAUT. In: Proceedings of the 2025 International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS); 2025 Feb 3–4; Bandung, Indonesia. p. 1–6.
27. Apau R, Titis E, Lallie HS. Towards a better understanding of mobile banking app adoption and use: integrating security, risk, and trust into UTAUT2. Computers. 2025;14(4):144. doi:10.3390/computers14040144.
28. Desai B, Patil K, Mehta I, Patil A. Explainable AI in cybersecurity: a comprehensive framework for enhancing transparency, trust, and human-AI collaboration. In: Proceedings of the 2024 International Seminar on Application for Technology of Information and Communication (iSemantic); 2024 Sep 21–22; Semarang, Indonesia. p. 135–50.
29. Fernández-Martínez C, Bikos A, Verikoukis C, Siddiqui S. Trusted access to 6G testbeds through a security intent-driven software-defined perimeter framework. In: Proceedings of the 2024 IEEE 29th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD); 2024 Oct 21–23; Athens, Greece. p. 1–6.
30. Ayushi, Dubey V, Galhotra B. Regulatory compliance and user trust: balancing innovation and security in AI-driven online payment systems. In: Proceedings of the 2024 4th International Conference on Sustainable Expert Systems (ICSES); 2024 Oct 15–17; Kaski, Nepal. p. 442–8.
31. Cao H, Yang L, Garg S, Alrashoud M, Guizani M. Softwarized resource allocation of tailored services with zero security trust in 6G networks. IEEE Wirel Commun. 2024;31(2):58–65. doi:10.1109/mwc.001.2300383.
32. Heydari V, Nyarko K. Fairness in machine learning for cybersecurity: enhancing trust through feature importance and SHAP analysis. In: Proceedings of the 2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME); 2024 Nov 4–6; Male, Maldives. p. 1–6.
33. Kaul P, Mane A, Mendonca C, Maurya V. MalwareGuard: Trust score prediction system. In: Proceedings of the 2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA); 2024 Dec 20–21; Vijayapura, India. p. 1–6.
34. Kumar D, Pramod Pawar P, Kumar Meesala M, Kumar Pareek P, Reddy Addula S, Shwetha KS. Trustworthy IoT infrastructures: privacy-preserving federated learning with efficient secure aggregation for cybersecurity. In: Proceedings of the 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS); 2024 Nov 22–23; Kalaburagi, India. p. 1–8.
35. Nodehi NR, Berisha F, Da Silva L, Pourzolfaghar Z, Helfert M. Integrating cybersecurity, data sovereignty and trustworthiness in agri-data sharing environments: a conceptual framework. In: Proceedings of the 2024 Cyber Research Conference—Ireland (Cyber-RCI); 2024 Nov 25; Carlow, Ireland. p. 1–8.
36. Tam C, Balau M, Oliveira T. What influences people's adoption of cognitive cybersecurity? Int J Hum Comput Interact. 2024;40(23):8295–312. doi:10.1080/10447318.2023.2279411.
37. Zolanvari M, Yang Z, Khan K, Jain R, Meskin N. TRUST XAI: model-agnostic explanations for AI with a case study on IIoT security. IEEE Internet Things J. 2023;10(4):2967–78.
38. Ramya S, Azad SMAK. Dynamic trust-based process control system for enhanced industrial security. In: Proceedings of the 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA); 2023 Dec 21–23; Bengaluru, India. p. 596–602.
39. Han J, Cho A. Practical in-vehicle security architecture based on trust anchors. In: Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring); 2023 Jun 20–23; Florence, Italy. p. 1–3.

40. Munir MS, Shetty S, Rawat DB. Trustworthy artificial intelligence framework for proactive detection and risk explanation of cyber attacks in smart grid. In: Proceedings of the 2023 Winter Simulation Conference (WSC); 2023 Dec 10–13; San Antonio, TX, USA. p. 636–47.
41. Khan A, Sharma I. TrustDroid: enhancing trust and privacy for Android mobile phones by preventing riskware and SMS malware. In: Proceedings of the 2023 2nd International Conference on Futuristic Technologies (INCOFT); 2023 Nov 24–26; Belagavi, India. p. 1–6.
42. Siemers B, Fischer L, Lehnhoff S. A trust model in control systems to enhance and support cybersecurity. In: Proceedings of the 2022 IEEE 7th International Energy Conference (ENERGYCON); 2022 May 9–12; Riga, Latvia. p. 1–6.
43. Dong Z, Qin Y, Li Z, Wang H. A security and trust protection framework for open CNC production line. In: Proceedings of the 2022 China Automation Congress (CAC); 2022 Nov 25–27; Xiamen, China. p. 6481–6.
44. Neises J, Evangelatos S, Soldatos J, Walloschke T, Moldovan G, Eikerling H, et al. Trustworthy human-machine assistance by dynamic process security monitoring in industrial environments. In: Sibalija T, Davim JP, editors. Soft computing in smart manufacturing: solutions toward Industry 5.0. Berlin/Boston: De Gruyter; 2021. p. 55–82. doi:10.1515/9783110693225-002.
45. Kadena E, Salvador LCR, Rajnai Z. Behavioral Biometrics for more (dis) trust and security. In: Proceedings of the 2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI); 2022 May 25–28; Timisoara, Romania. p. 000081–6.
46. Greaves DJ, Sekhon H, Garcia-Perez A. Pathways to trust: a model of information security and trust formation in socio-technical environments structured abstract. In: Proceedings of the Competitive Advantage in the Digital Economy (CADE 2021); 2021 Jun 2–3; Online Conference. p. 199–203.
47. Khan WZ, Khurram Khan M, Arshad QUA, Malik H, Almuhtadi J. Digital labels: influencing consumers trust and raising cybersecurity awareness for adopting autonomous vehicles. In: Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE); 2021 Jan 10–12; Las Vegas, NV, USA. p. 1–4.
48. Deljoo A, Koning R, van Engers T, Gommans L, de Laat C. Managing effective collaboration in cybersecurity alliances using social computational trust. In: Proceedings of the 2019 3rd Cyber Security in Networking Conference (CSNet); 2019 Oct 23–25; Quito, Ecuador. p. 50–7.
49. Ghernaouti S, Cellier L, Wanner B. Information sharing in cybersecurity: enhancing security, trust and privacy by capacity building. In: Proceedings of the 2019 3rd Cyber Security in Networking Conference (CSNet); 2019 Oct 23–25; Quito, Ecuador. p. 58–62.
50. Cellier L, Ghernaouti S. An interdisciplinary approach for security, privacy and trust in the electronic medical record: a pragmatic legal perspective. In: Proceedings of the 2019 IEEE International Conference on E-Health Networking, Application & Services (HealthCom); 2019 Oct 14–16; Bogota, Colombia. p. 1–6.
51. Markakis E, Nikoloudakis Y, Pallis E, Manso M. Security assessment as a service cross-layered system for the adoption of digital, personalised and trusted healthcare. In: Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT); 2019 Apr 15–18; Limerick, Ireland. p. 91–4.
52. Kharvi PL. A design science framework for measuring trust and security in the metaverse space: a holistic approach to digital trustworthiness [dissertation]. Arlington, VA, USA: Marymount University; 2025.
53. Alhasan I. Human factors in cybersecurity: a cross-cultural study on trust [Ph.D. thesis]. West Lafayette, IN, USA: Purdue University; 2023.
54. Kim DY. Trusted compliance enforcement framework for large volume and high velocity data [Ph.D. thesis]. Baltimore, MD, USA: University of Maryland, Baltimore County; 2023.
55. Osunji O. Government's role in building trust and confidence in the internet: a case study of Uganda in the implementation of cybersecurity capacity maturity model for nations [dissertation]. Arlington, VA, USA: Marymount University; 2022.
56. Yeoh W, Liu M, Shore M, Jiang F. Zero trust cybersecurity: critical success factors and a maturity assessment framework. Comput Secur. 2023;133:103412.
57. Pakaja F. Overconfidence bias measures and herd behavior on information system security. J Comput Inf Syst. 2025:1–17. doi:10.1080/08874417.2025.2507707.

58. Palbar Misas JD, Hopcraft R, Tam K, Jones K. Future of maritime autonomy: cybersecurity, trust and mariner's situational awareness. J Mar Eng Technol. 2024;23(3):224–35. doi:10.1080/20464177.2024.2330176.
59. Bajwa IA, Ahmad S, Mahmud M, Bajwa FA. The impact of cyberattacks awareness on customers' trust and commitment: an empirical evidence from the Pakistani banking sector. Inf Comput Secur. 2023;31(5):635–54. doi:10.1108/ICS-11-2022-0179.
60. Itodo C, Ozer M. Multivocal literature review on zero-trust security implementation. Comput Secur. 2024;141:103827. doi:10.1016/j.cose.2024.103827.
61. van der Schyff K, Flowerday S. The mediating role of perceived risks and benefits when self-disclosing: a study of social media trust and FoMO. Comput Secur. 2023;126:103071. doi:10.1016/j.cose.2022.103071.
62. Frank M, Jaeger L, Manuel Ranft L. Using contextual factors to predict information security overconfidence: a machine learning approach. Comput Secur. 2023;125:103046. doi:10.1016/j.cose.2022.103046.
63. Collett R. Understanding cybersecurity capacity building and its relationship to norms and confidence building measures. J Cyber Policy. 2021;6(3):298–317. doi:10.1080/23738871.2021.1948582.
64. Butavicius M, Parsons K, Lillie M, McCormac A, Pattinson M, Calic D. When believing in technology leads to poor cyber security: development of a trust in technical controls scale. Comput Secur. 2020;98:102020.
65. Sumithra M, Kiruthika S, Nithya S, Poornima B, Dharanya S. Enhancement of cloud user data access security entrusted to AI face recognition techniques. J Cogn Hum Comput Interact. 2022;2(2):60–4. doi:10.54216/jchci.020204.
66. Owen M, Flowerday SV, van der Schyff K. Optimism bias in susceptibility to phishing attacks: an empirical study. Inf Comput Secur. 2024;32(5):656–75. doi:10.1108/ICS-02-2023-0023.
67. Vyhmeister E, Castane GG. TAI-PRM: trustworthy AI—project risk management framework towards Industry 5.0. AI Ethics. 2025;5(2):819–39. doi:10.1007/s43681-023-00417-y.
68. Razali NAM, Ishak KK, Md Saad NJA, Wook M, Ramli S. The assessment of trust in information security using Kansei. In: Shoji H, Koyama S, Kato T, Muramatsu K, Yamanaka T, Lévy P et al., editors. Proceedings of the 8th International Conference on Kansei Engineering and Emotion Research. Singapore: Springer; 2020. p. 194–202.
69. Pigola A, de Souza Meirelles F. Unraveling trust management in cybersecurity: insights from a systematic literature review. Inf Technol Manag. 2026;27(1):71–93. doi:10.1007/s10799-024-00438-x.
70. Chen Y, Zahedi FM, Abbasi A, Dobolyi D. Trust calibration of automated security IT artifacts: a multi-domain study of phishing-website detection tools. Inf Manag. 2021;58(1):103394. doi:10.1016/j.im.2020.103394.
71. Wischnewski M, Krämer N, Müller E. Measuring and understanding trust calibrations for automated systems: a survey of the state-of-the-art and future directions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; 2023 Apr 23–28; Hamburg, Germany. p. 1–16.
72. Yazdanmehr A, Wang J, Yang Z. Peers matter: the moderating role of social influence on information security policy compliance. Inf Syst J. 2020;30:791–844.
73. Zhuo S, Biddle R, Koh YS, Lottridge D, Russello G. SoK: human-centered phishing susceptibility. ACM Trans Priv Secur. 2023;26(3):1–27. doi:10.1145/3575797.
74. Hagen RA, Øverlier L, Helkala K. Human factors in AI-driven cybersecurity: cognitive biases and trust issues. Digit Threat Res Pract. 2025;6(4):1–20. doi:10.1145/3759260.
75. Klingbeil A, Grützner C, Schreck P. Trust and reliance on AI—an experimental study on the extent and costs of overreliance on AI. Comput Hum Behav. 2024;160:108352. doi:10.1016/j.chb.2024.108352.
76. Chaudhary S. Driving behaviour change with cybersecurity awareness. Comput Secur. 2024;142:103858. doi:10.1016/j.cose.2024.103858.
77. Okamura K, Yamada S. Adaptive trust calibration for human-AI collaboration. PLoS One. 2020;15(2):e0229132. doi:10.1371/journal.pone.0229132.
78. Lee CS, Kim D. Pathways to cybersecurity awareness and protection behaviors in South Korea. J Comput Inf Syst. 2023;63(1):94–106. doi:10.1080/08874417.2022.2031347.

79. Taib R, Yu K, Berkovsky S, Wiggins M, Bayl-Smith P. Social engineering and organisational dependencies in phishing attacks. In: Human-Computer Interaction—INTERACT 2019. New York, NY, USA: The Association for Computing Machinery (ACM); 2019. p. 564–84.
80. Garba J, Kaur J, Ibrahim ENM. Design of a conceptual framework for cybersecurity culture amongst online banking users in Nigeria. Niger J Technol. 2023;42(3):399–405. doi:10.4314/njt.v42i3.13.