



Kent Academic Repository

Spicer, Mark and Nurse, Jason R. C. (2026) *What makes people susceptible to phishing attacks: A systematic literature review and future research agenda.* Computers & Security . ISSN 0167-4048.

Downloaded from

<https://kar.kent.ac.uk/115870/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.cose.2026.105096>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

What makes people susceptible to phishing attacks: A systematic literature review and future research agenda

Mark Spicer and Jason R. C. Nurse

School of Computing and Institute of Cyber Security for Society (iCSS),
University of Kent, UK

Highlights

- Synthesises 54 studies on individual/environmental factors in phishing susceptibility.
- Existing evidence dominated by educational/Western settings limiting generalisability.
- Opinion is divided in many cases as to the key influencing factors.
- Interaction between human and environmental factors is an area that requires specific attention in future research.
- More large-scale controlled studies conducted in real-world industry environments required.

Abstract

Despite significant advances in both cybersecurity technology and anti-phishing research, phishing remains the most common cause of breaches. This Systematic Literature Review synthesises existing evidence on why people fall for phishing emails, examining the individual and environmental factors associated with phishing susceptibility and the effectiveness of organisational controls. Following the PRISMA methodology, 4,402 articles were identified, of which 54 met the inclusion criteria and formed the basis of analysis. This review makes three contributions. First, it produces a systematic evidence verdict on each factor associated with phishing susceptibility, separating those with strong consensus support from those that are contested, robust nulls, or supported by only thin evidence. Second, it develops a conceptual framework integrating individual factors, environmental factors and the intention-behaviour gap as three contributors to the decision process. Third, it identifies a structural mismatch between current organisational controls, which predominantly target individual knowledge through awareness training and the multi-factor structure of susceptibility, which requires intervention across all three contributors. Practical implications for organisational cybersecurity practice and a research agenda spanning empirical, ethical and AI-enabled-phishing dimensions are developed. It also confirms the need for more large-scale research in a real-world environment to understand individual and environmental factors that cause people to be susceptible to phishing emails.¹

Keywords: Phishing, social engineering, human factors, individual factors, psychological factors, environmental factors, demographic factors

¹ This is the author accepted manuscript of the paper: “What makes people susceptible to phishing attacks: A systematic literature review and future research agenda” by Mark Spicer and Jason R. C. Nurse due to appear in Computers and Security Journal (2026), <https://doi.org/10.1016/j.cose.2026.105096>

1. Introduction

We live in a world in which technology surrounds us and in which the adverse consequences of cyber-attacks have grown exponentially (Steinberg, 2022). Phishing emails represent the most prevalent and disruptive form of cyber-attack currently facing organisations, experienced by 85% of businesses according to the most recent UK government survey (GOV.UK, 2025).

Phishing attacks occur when adversaries send fraudulent emails containing malicious links designed to trick recipients into revealing sensitive information (NCSC, 2024). Despite significant investment in both technical defences and user awareness training, phishing continues to be the primary vector for data breaches and organisational compromise (Petric and Just, 2025; Owoh and Haig, 2025). The persistence of the problem despite decades of awareness efforts points to a fundamental gap in understanding: not simply how many people fall for phishing emails, but why.

One specific instance of this is the intention-behaviour gap, a phenomenon in which stated intentions to act in a particular way do not reliably translate into observed behaviour. In the phishing context, employees who report awareness of phishing risks and express secure intentions during training nevertheless click on phishing emails under operational conditions. As this review shows, the intention-behaviour gap emerges as a central failure mode in current understanding of phishing susceptibility and is integrated as one of the three contributors in the conceptual framework developed in Section 4.

A substantial body of research already exists on individual susceptibility to phishing, including several systematic reviews and structured literature surveys that examined aspects of this problem. Desolda et al. (2022) provided a broad review of human factors in phishing, cataloguing vulnerabilities such as lack of knowledge, awareness, and complacency. Das et al. (2020) re-examined phishing research from a security standpoint, while Das et al. (2022) combined a systematic review of ACM Digital Library publications with a user study of high school students and educators. Baki and Verma (2023) provided a meta-analysis of sixteen years of phishing user studies, reporting consolidated effects for demographic factors and training-based interventions. Zhuo et al. (2023) used temporal analysis, identifying three time periods that frame phishing susceptibility, and Tornblad et al. (2021) catalogued thirty-two known predictors of phishing susceptibility.

The present review builds on this body of work but differs from it in three specific ways. First, it provides a unified synthesis across psychological, demographic, behavioural, and environmental factors, alongside the organisational controls intended to mitigate them, rather than addressing any one of these dimensions in isolation. Second, it critically evaluates the methodological quality and real-world applicability of existing large-scale studies, identifying a consistent pattern of limited researcher control over content, context, and persuasion techniques of phishing stimuli. This is a methodological constraint that limits causal inference, acknowledged by many of the original authors as a limitation of their work. Third, it situates findings within the context of organisational controls specifically, asking whether the psychological and environmental factors identified in the literature are reflected in the design of

awareness training and other interventions, and proposes a structured future research agenda grounded in the gaps identified.

The purpose of this research is to synthesise existing knowledge on the individual and environmental factors that cause people to be susceptible to phishing emails and to provide a foundation for future research. This is achieved through a comprehensive SLR of peer-reviewed studies published predominantly over the last ten years. In doing so, the review aims to answer three research questions (RQs):

RQ1. Which individual factors including psychological (personality traits, cognitive styles), demographic (age, gender) and behavioural (e.g., impulsivity, intended versus actual behaviour) characteristics are associated with increased susceptibility to phishing emails and to what extent do these factors predict phishing-related behaviours such as link clicking or credential disclosure?

RQ2. How do environmental factors including workload, time pressure, time of day and the home versus work context influence the likelihood of users engaging with phishing emails?

RQ3. How effective are organisational controls including Cybersecurity Awareness and Training (CSAT) and phishing simulations in reducing user susceptibility to phishing and to what extent do they account for the individual and environmental factors identified in RQs 1 and 2?

Whilst the search strategy employed the term 'human' to ensure broad retrieval of relevant literature, consistent with its widespread use in the phishing susceptibility literature, this paper adopts the more precise term 'individual factors' when characterising the findings. This distinction is important as 'human factors' is a distinct and established academic field concerned with ergonomics and human-system interaction (HFES, 2024). The factors examined in this research relate specifically to characteristics of individuals (factors about humans) their psychology, demographics and behaviour, rather than to human-systems design more broadly.

This review makes three contributions to the literature on phishing susceptibility. First, the review provides a systematic evidence verdict on each individual, environmental and organisational factor identified across the corpus. It distinguishes factors with strong empirical consensus from those that are contested, robust nulls or supported by thin evidence. The verdict is produced through formal vote-counting (Section 2.8) combined with a quality assessment of each included study (Section 2.7). Second, the review develops a conceptual framework (Section 4) that integrates the three categories of contributor into a single model centred on the intention-behaviour gap. This draws on dual-process theory and habit-and-automaticity to explain why stated security intentions do not translate into observed behaviour. Third, the framework reveals a structural mismatch between where organisational controls currently focus (predominantly on individual knowledge through awareness training) and the multi-factor structure of susceptibility, with implications for security practice developed in Section 4.5.2.

The remainder of this paper is structured as follows. Section 2 explains the research method, including the application of the PRISMA framework. Section 3 presents the results of the SLR,

organised by research question. Section 4 provides a discussion of the key findings, their implications, a structured future research agenda and the limitations of the review. Finally, Section 5 provides a conclusion.

2 Research method

To ensure methodological consistency and replicability, this SLR was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, which provides a standard methodology that contributes to the quality assurance of the revision and replication process (Tedja et al., 2024).

2.1 Search strategy

The anticipated benefit of a SLR is to conduct it in a way that it can produce repeatable results (Kitchenham et al., 2011) and is now widely used in cybersecurity systematic reviews such as Patterson et al., 2023; Uchendu et al., 2021 and Naqvi et al., 2023. The choice of keywords for the search is imperative for the selection of articles for inclusion in a SLR, as these terms are the basis for access to relevant articles (Salloum et al., 2022).

The search terms selected were grounded in the aims of the research and its defined research questions, thus terms such as "social engineering", "phishing", and "human" were used. The term social engineering was used as it is the act of manipulating human beings most commonly with the use of psychological persuasion (Washo, 2021). The term phishing was used as it has become one of the most prominent forms of social engineering attacks faced by internet users (Carroll et al., 2022). To ensure the individual human aspects of phishing susceptibility were included, references to psychology were also included. The latter term was changed to include as many possible versions of psychology so the wildcard "psychol*" was used. Keeping the search terms relatively broad ensured the initial retrieval was wide, retrieving 4,402 articles, whilst the subsequent eligibility criteria were used to focus the selection on relevant studies.

The terms were further refined depending on the syntaxes required by each digital library searched. The final keywords used were "social engineering", "phishing", "human" and "psych(ology)". These terms are thought to be reasonably well-defined terms suitable for searching. Thus, the authors' approach allowed the authors to balance being too narrow and missing publications or being too broad and generating unwieldy initial results, following the keyword-refinement approach used by Patterson et al. (2023).

The authors then built sophisticated search strings using Boolean ANDs and ORs as suggested by Kitchenham and Charters (2007). Keeping the search terms broad meant 4,402 articles were included in the initial review. For repeatability and transparency, the exact search terms used are shown in Table 1. The searches took place on the 20th of December 2025.

Digital Library	Field	Search term used
IEEE Xplore	Abstract, title, keywords	((("Abstract": "social engineering" OR "Abstract": "phishing") AND ("Abstract": "human" OR "Abstract": "psychol*")) OR (("Document Title": "social engineering" OR "Document Title": "phishing") AND ("Document Title": "human" OR "Document Title": "psychol*")) OR (("Author Keywords": "social engineering" OR "Author Keywords": "phishing") AND ("Author Keywords": "human" OR "Author Keywords": "psychol*"))))
The ACM Guide to Computing Literature	Abstract, title, keywords	(Title: ("social engineering" OR "phishing") AND Title: ("human" OR "psychol*")) OR (Abstract: ("social engineering" OR "phishing") AND Abstract: ("human" OR "psychol*")) OR (Keyword: ("social engineering" OR "phishing") AND Keyword: ("human" OR "psychol*"))
Elsevier	Abstract, title, keywords	Title, abstract, keywords: ('social engineering' OR 'phishing' AND ('human' OR 'psychological'))
Scopus	Abstract, title, keywords	("social engineering" OR "phishing") AND (<u>"human"</u> OR psychol*)
Web of science	Abstract, title, keywords	("social engineering" OR "phishing") AND ("human" or "psychol*")
ProQuest One	Anywhere except full text	noft(("social engineering" OR "phishing") AND ("human" OR "psychol*"))
APA PsycInfo	Abstract, title, keywords	("social engineering" OR "phishing") AND ("human" or "psychol*")

Table 1 - Literature search queries by library

2.2 Eligibility criteria

The articles found in the search were reviewed using the criteria shown in Table 2.

Screening Criteria		
Inclusion Criteria	IC1	Articles published between 2016 and 2025 (inclusive), applied at the library search stage.
	IC2	Peer-reviewed articles published in English, applied at the library search stage and during manual review.
	IC3	Papers addressing individual (psychological, demographic, and behavioural) factors and/or environmental factors of phishing susceptibility, in alignment with the research questions.
	IC4	Research with datasets covering more than 50 participants or 200 emails, to ensure a minimum level of statistical viability.
	IC5	Primary or secondary research based on real-world studies undertaken in a workplace or educational environment.

Exclusion Criteria	EC1	Duplicates and repeated studies.
	EC2	Non-academic and non-peer-reviewed articles, including books and theses or dissertations (except those at PhD level).
	EC3	Studies for which the full text was unavailable.
	EC4	Studies not relevant to the research questions.

Table 2 – Inclusion/Exclusion criteria

For IC4, research with datasets covering more than 50 participants or 200 emails reflects minimum thresholds recommended for behavioural statistical analysis (Hair et al., 2018; Guilford, 1954). Two distinct thresholds were used as they apply to different study designs: (the participant threshold applies to survey /and interview designs, and the email threshold applies to simulation-based designs). The 200-email minimum applies Guilford's (1954) statistical threshold of 200 paired observations for Pearson correlation analysis to phishing-simulation studies, where individual emails constitute the unit of analysis. The 50-participant minimum follows Hair et al. (2018), as cited by Memon et al. (2020), who establish 50 observations as the lower bound for exploratory factor analysis and regression-based behavioural research, with 100 participants recommended where feasible. These thresholds were applied as minimum bars for statistical defensibility rather than as quality criteria; studies meeting the thresholds were subsequently differentiated by the formal quality assessment described in Section 2.7.

As each title, keywords and abstract were reviewed, they were selected if they satisfied the inclusion and exclusion criterion.

For the purposes of this review, 'real-world' research studies are defined as falling within one of three categories: (1) phishing simulations delivered to participants' inboxes without prior notification in which responses (clicking, ignoring, or reporting) are observed; (2) the use of historic datasets derived from actual phishing emails received by individuals; or (3) structured interviews, surveys, controlled experiments in which participants provide psychological, demographic or behavioural data relating to phishing susceptibility. It is acknowledged that category (3) captures a wide range of study designs, however all are grounded in real-participant responses rather than purely hypothetical or modelled data.

2.3 Information Sources

Prior to conducting the search, an initial scoping exercise was carried out to identify the databases most frequently used in related SLRs. Seven peer-reviewed databases were ultimately selected: IEEE Xplore, the ACM Digital Library, Science Direct, Scopus, Web of Science, ProQuest One and APA PsycINFO. Scopus and Web of Science were included as databases of databases, to maximise breadth of coverage. ProQuest One and PsycINFO were included because of their coverage of psychological and behavioural research, which is directly relevant to the individual factors examined in this review. Springer was considered but excluded because it did not allow searching to be restricted to abstract, title, and keywords, which was a requirement for consistency across databases.

The search terms were required to appear in the abstract, title, or keywords of each article. The exact search strings applied to each database are shown in Table 2.

2.4 Search period

Previous research into scholarly interest in phishing-related research found only a modest increase in publications between 2007 and 2015 and then from 2016 onwards a notable surge occurred according to Scopus (Ghazali et al., 2025). Thus, a date range of 2016 to 2025 was felt sufficient and appropriate although this was later partly extended as discussed in Section 2.5.

2.5 Study selection

The aim of this SLR was to find as many primary studies relating to the research question as possible using an unbiased search strategy and thus SLRs were excluded from the research. The rigour of the search process distinguishes SLRs from traditional reviews (Kitchenham and Charters, 2007). Preliminary searches were carried out, aimed at assessing the volume of potentially relevant studies. The search process was carried out 3 times and iteratively refined to ensure accuracy. In keeping with the PRISMA framework, the exact final search terms used for each digital catalogue are shown in Table 2. Each database has its own syntax requirements.

The search process retrieved 4,402 peer-reviewed articles written in English (IC2) and published between 2016 and 2025 (IC1) with the bulk being from Scopus. These were loaded into a bibliography tool (Zotero) where it was found a significant number (1477) were duplicates which were then removed (EC1) leaving 2925 papers.

A review of the titles, keywords and abstracts of the remaining 2925 papers was then carried out to assess their relevance meaning 2744 were excluded (exclusion criteria EC2 and EC4) leaving 181 papers. A full text review took place to satisfy criteria IC3, IC4 and IC5 plus a full text review (EC3) resulting in 145 papers being excluded leaving 36 papers.

The reference sections of these articles were reviewed for further relevant articles and each were reviewed using the inclusion criteria, a process called reverse snowballing (Desolda et al., 2022). As many of the articles reviewed in this way were cited many times and could therefore be key publications, no date criteria was applied. This method followed the approach by Desolda et al., who argued this mitigates possible limitations of the inclusion and exclusion criteria, permitting authors to retrieve relevant papers published outside of inclusion criteria date range. Thus, an additional 15 relevant papers were included. The authors repeated the review process of these new articles and a further 2 relevant papers were found (17 in total through reference searches). As noted, the searches took place on the 20th December 2025 but to ensure more recent relevant articles were not missed, a Google Scholar alert was put in place. This highlighted one additional, recent article to include. A review of the reference section of this article found no further relevant articles leaving the number of articles reviewed as 54. Figure 1 illustrates the PRISMA flow diagram for this process.

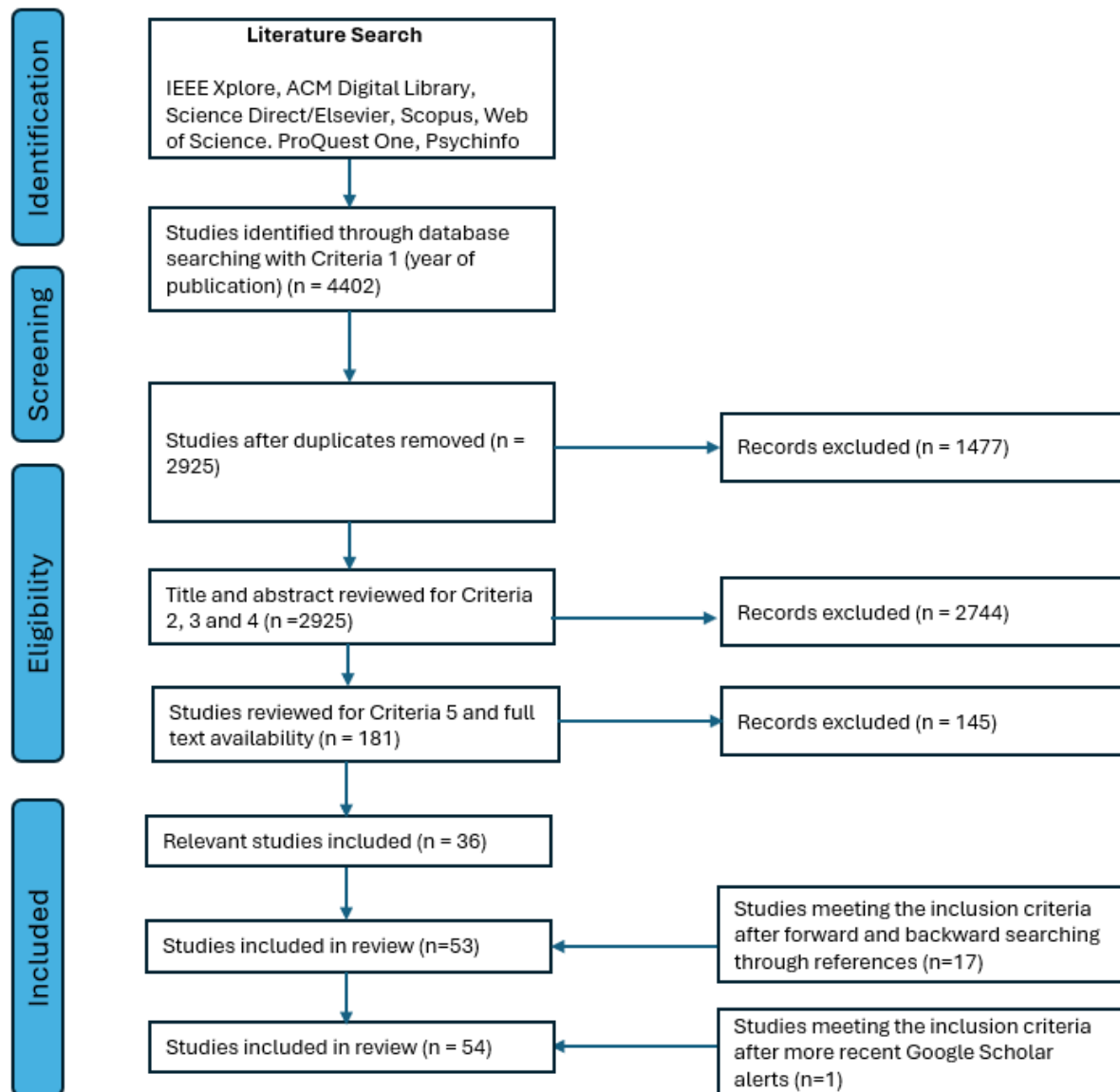


Figure 1 - PRISMA workflow

2.6 Data collection

Data were collected in accordance with PRISMA requirements focusing on previous research relevant to the research questions. For example, the authors extracted information on research focused on the individual (psychological, demographic and behavioural) factors associated with increased susceptibility (RQ1). Also, information regarding environmental (e.g., time of email delivery, workload, environment (home versus work)) factors (RQ2) was extracted. Finally, we extracted information on the effects of organisational controls (RQ3). All data were reviewed and extracted by a single reviewer using the inclusion/exclusion criteria. During this process no assumption was made that the papers selected exclusively address specific issues and conversely many papers selected addressed more than one of the research questions.

2.7 Quality assessment of included studies

PRISMA 2020 emphasises transparent reporting of risk-of-bias assessments when they are conducted, but it does not mandate that systematic reviews must include a formal appraisal of study quality. In our review we distinguish findings supported by methodologically stronger evidence from those resting on weaker evidence, particularly in relation to risk of bias, dataset realism, and sample strength. Each of the 54 included studies was appraised using a six-criterion quality assessment framework. Three criteria were adapted from the established quality assessment checklist of Dybå and Dingsøyr (2008), itself developed within the Evidence-Based Software Engineering tradition (Kitchenham and Charters, 2007), and three field-specific criteria added in response to Reviewer 3's feedback.

The three general criteria that were adapted are: (Q1) clarity of aims and research questions; (Q2) appropriateness of the research design and (Q6) clarity of findings and how they are supported by the reported data. The three field-specific criteria are (Q3) risk of bias, capturing self-report bias, awareness bias and sampling bias; (Q4) dataset realism or ecological validity, distinguishing laboratory studies, third-party dataset analyses and real-world operational simulations and (Q5) sample strength, assessing whether the sample size and composition adequately support the inferences drawn. Each criterion is scored on a three-point scale: 0 (criterion not met), 0.5 (partially met) or 1 (fully met), giving a maximum total score of 6.

To minimise the subjectivity inherent in quality assessment, scoring rules were operationalised against objective study characteristics extractable from the published papers rather than relying on reviewer judgement alone. Risk of bias was assessed against author count and study type (real-world simulations carrying lower bias than self-report studies); dataset realism against study type (real-world simulations highest, third-party dataset analyses intermediate, laboratory experiments lowest); sample strength against tiered thresholds appropriate to each study type. Clarity of findings was assessed against the number of times each paper was cited within the review's analytical discussion and the number of topics it contributed to in the synthesis. Studies identified as seminal to the field (Sheng et al., 2010; Steves et al., 2020; Vishwanath et al., 2011, 2018; Oliveira et al., 2017) were protected from sample-based penalties in recognition of their foundational contribution. The full scoring framework is documented in Appendix A, the scoring criteria are documented in Appendix B.

Scoring was conducted by the first author with independent spot-checks by the second author. Single-reviewer bias is acknowledged as a residual limitation in Section 4.7. The findings of the quality assessment, including the distribution of studies across quality tiers, are reported in Section 3.4 alongside the synthesis of evidence by research question.

2.8 Synthesis method and approach

The 54 included studies were synthesised using a vote-counting approach (Bushman and Wang, 2009). For each factor examined in the review, the direction of finding reported by each study was classified as significant (✓), null (X), or mixed or partial (~) and the resulting counts were tallied across the corpus. The results of applying this approach are presented in Section 3 (Table 3). Classification followed the original authors' own statistical-significance judgements rather than re-classifying findings, preserving consistency with the published literature.

Vote-counting was chosen over meta-analysis because the methodological heterogeneity of the corpus makes the calculation of comparable effect sizes inappropriate. The 54 included studies encompass real-world phishing simulations (32 studies), third-party dataset analyses (6 studies) and lab experiments or structured interviews (16 studies). These have widely different outcome measures including raw click rate, self-reported intention, detection accuracy and

aggregated susceptibility scores. Pooling effect sizes across such heterogeneous designs and measures would obscure rather than reveal the substantive patterns in the literature.

Two limitations of vote-counting should be acknowledged. First, the approach cannot statistically weight studies by sample size or methodological quality, meaning a single small lab study contributes the same vote as a large operational simulation. The quality assessment introduced in Section 2.7 partially mitigates this. Second, vote-counting is sensitive to the categorical thresholds used to classify findings. Both limitations are acknowledged in Section 4.7. The results of applying this approach to the corpus are reported in Section 3 (Table 8) which represents a study-level direction of finding for each factor; the basis for the vote-counting summarised in Table 6.

3. Results

This section presents the findings of the review. Table 3 below presents the synthesis matrix across the 54 included studies, with each cell recording whether the study reported a significant effect (✓), a null result (X) or a mixed or partial result (~) for each factor. Empty cells indicate the study did not address that factor.

Section 3.1 reports the characteristics of the research corpus, including the distribution of methodological quality scores from Section 2.7. Sections 3.2 through 3.4 then report findings organised by research question (RQ1: individual factors; RQ2: environmental factors; RQ3: organisational controls).

First Author	RQ1 — Individual factors														RQ2 — Environmental factors						RQ3 — Organisational controls					
	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Altruism	Personality traits: no effect found	Impulsivity	Self-efficacy	Cognitive processing	Intention– behaviour gap	Urgency cue	Authority cue	Social proof/Scarcity cue	Liking/Deception cue	Gender	Age	Workload/Time pressure	Email volume	Time of email delivery (proposed)	Insider spoofing/Source of email	Work vs home context	Env × Individual interaction noted	Overall finding	Tailored/targeted training more effective	Generic training insufficient/one-size-fits-all
Abroshan																✓										
Albladi									✓																	
Alwanain																?								✓	✓	
Anawar	✓	✓	?	?	?											?								✓		
Bakhshi												✓	✓		✓			✓		?		✓		✓		
Bayl-Smith	✓	?		?					✓					✓				✓						✓		
Bollens												✓							✓							
Bona												✓	?			✗						✓	✓	?		
Braun																								✗		
Champa	✓											✓	✓	✓										✓	✓	
Chrysanthou																		✓						✓		
Das																	✓							✓		
Eftimie	✓	✓		✓												✗				?				?		
Ferreira												?	?		✓											
Frank																✓	✓							✓	✓	
Gallo												?				?	?	✓	✓					✓		
George											✓															
Gordon																		✓								

Gratian	✓	~	✓	✓	~											✓												
Hanus																	✓							✓	✓			
Harrison										~														X				
Hassandoust										✓						✓								✓				
Jalali										✓								✓						✓				
Jones								✓										✓										
Koddebusch											✓	✓				X												
Kumar	✓															X	✓											
Lain (2021)																								X				
Lain (2022)																		✓						X				
Lawson	✓	✓										✓	✓	✓		X								~				
Longtchi								✓																				
Luga																✓								X				
McElwee																								✓	✓	X		
Moody	X	X	X	X	X		✓									X												
Musava										X														✓	✓			
O'Connor																	✓											
Oliveira										✓			✓			✓	✓											
Oner																✓	✓							✓	✓			
Parsons								✓								X								X		X		
Pinto																								✓	✓			
Sarno (2020)																	✓		~									
Sarno (2022)																		✓	✓					✓				
Scott																								X		X		
Sheng																✓	✓							✓				
Smith											✓	✓																

Steves												✓	✓											✓		
Sutter																							✓	✓		
Taib															~	✓										
Toth						✓					✓		✓						✓			✓	✓	✓		
Vishwanath (2011)									X	✓		✓						✓	✓							
Vishwanath (2018)										✓																
Wang												✓	✓													
Williams	✓											✓	✓					✓			✓	✓		X		X
Yang															✓								✓	✓		
Zielinska												✓	✓	~												

Total - Significant (✓)	8	3	1	2	0	1	1	3	2	2	5	12	10	4	2	9	9	12	4	0	2	3	3	21	11	0	130
Total - not significant (X)	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0	7	0	0	0	2	0	0	0	8	0	4	28
Total - partial/mixed (~)	0	2	1	2	2	0	0	0	0	0	1	2	2	1	0	4	1	0	1	0	0	0	0	3	0	0	22
Total	9	6	3	5	3	1	1	3	3	3	6	14	12	5	2	20	10	12	5	2	2	3	3	32	11	4	180

Table 3 – Synthesis Matrix of main research topics by paper and vote scoring totals

3.1 Characteristics of the corpus

3.1.1 Research and publication details

A complete list of articles found in our search is shown in Table 4 below.

Number	Study	Number	Study
1	Abroshan et al. (2021)	28	Champa et al. (2025)
2	Bakhshi (2017)	29	Das et al. (2020)
3	Bayl-Smith et al. (2021)	30	Jones (2016)
4	Bollens (2024)	31	Lawson et al. (2020)
5	Braun et al. (2025)	32	Luga et al. (2016)
6	Eftimie et al. (2022)	33	O'Connor et al. (2021)
7	Frank et al. (2022)	34	Oliveira et al. (2017)
8	Gordon et al. (2019)	35	Toth et al. (2025)
9	Hanus et al. (2022)	36	Albladi and Weir (2020)
10	Hassandoust et al. (2020)	37	Alwanain (2019)
11	Jalali et al. (2019)	38	Bona and Paci (2020)
12	Koddebusch (2022)	39	Chrysanthou et al. (2024)
13	Kumar et al. (2023)	40	Ferreira and Lanzini (2015)
14	Lain et al. (2021)	41	Gallo et al. (2024)
15	Longtchi et al. (2024)	42	George et al. (2020)
16	McElwee et al. (2018)	43	Gratian et al. (2018)
17	Musava et al. (2019)	44	Harrison et al. (2016)
18	Pinto et al. (2022)	45	Moody et al. (2017)
19	Scott et al. (2024)	46	Parsons et al. (2013)
20	Steves et al. (2020)	47	Sarno et al. (2020)
21	Sutter et al. (2022)	48	Sheng et al. (2010)
22	Taib et al. (2019)	49	Vishwanath et al. (2011)
23	Wang and Lutchkus (2023)	50	Vishwanath et al. (2018)
24	Williams et al. (2018)	51	Lain et al. (2022)
25	Yang et al. (2022)	52	Smith and Dupuis (2025)
26	Zielinska et al. (2016)	53	Oner et al. (2025)
27	Anawar et al. (2019)	54	Sarno et al. (2022)

Table 4 - A list of research articles reviewed

3.1.2 Year of publication

Nearly half of the articles found were published in a four-year period between 2019 and 2022 as in Figure 2 below. Articles published prior to 2016 were included following the reverse snowballing process and thus there are the lowest number of publications listed in those years. As discussed in Section 2.5 this approach mitigates possible limitations of the inclusion/exclusion criteria since it permits authors to retrieve relevant papers published outside of any inclusion criteria date range.

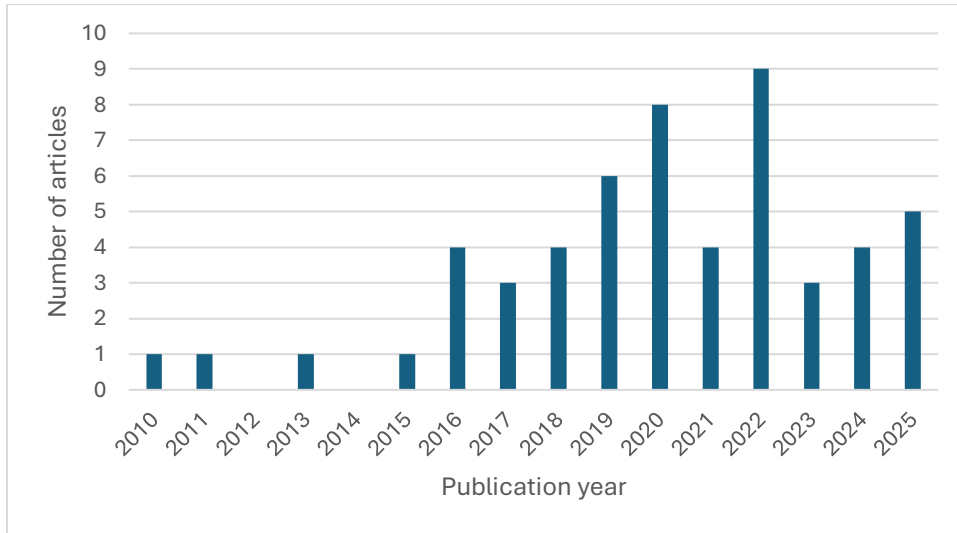


Figure 2 - Analysis of relevant articles by year of publication

3.1.3 Research sector

Twenty-five of the 54 studies were conducted in educational institutions such as universities or schools (Table 5). This concentration in educational settings has implications for the generalisability of findings to industry settings, as discussed in Section 4.

A full breakdown of the areas researched by industry type is shown in Table 5 below.

Sector/Research context	Number of articles
Education (Universities or schools)	25
Other Industry companies	9
Research volunteers	5
Library Email datasets	3
Public sector	3
IT-related companies	3
Financial Services	3
Hospitals	2
Unspecified	1

Table 5 - Sectors or research context. Note: 'Other industry companies' refers to papers where the authors described their setting as 'corporate' or 'pharmaceutical' without further specification type. Three entries denote research approach rather than sector.

3.1.4 Research topic

Each paper was categorised by research topic, individual factors (RQ1), environmental factors (RQ2) or organisational factors (RQ3) with a full breakdown shown in Table 3.

3.1.5 Context of research

Of the 36 studies specifying a country, 21 were carried out in the United States and a further 14 had no geographic context (online studies, library datasets, or where no country was specified). Henrich et al. (2010) argue that behavioural science routinely makes broad claims based on Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations, which are among the least representative populations for generalising about humans. Table 6 shows most reviewed studies were Western, consistent with Hasegawa et al.'s (2024) finding that 73% of usable privacy and security papers cover Western participants.

Category	Number of Articles
Specific country	
United States	21
Australia	3
United Kingdom	2
Canada	1
Fiji	1
Germany	1
Iran	1
Kenya	1
Malaysia	1
New Zealand	1
Portugal	1
Saudi Arabia	1
Switzerland	1
Geographic region (country unspecified)	
Europe (exact countries not specified)	3
Worldwide (multiple countries)	1
No defined geographic context	
Online (participants recruited online, no country specified)	5
Review of third-party datasets	5
Country not stated by authors	4

Table 6 - Geographic distribution of studies. Note: 'Online' indicates that participants recruited and/or data collected via online surveys or platforms, without a defined country context. 'Review of third-party datasets' indicates use of pre-existing historic phishing datasets.

3.1.6 Research type and frameworks used

The review identified 54 studies that could be grouped into three main research types: phishing simulations (32), third-party phishing datasets (6) and structured interviews or lab experiments (16). A breakdown of these is in Table 7 below.

Phishing simulations		Phishing datasets		Structured interviews or Lab experiments	
First author	No of Emails sent	Main author	No of Emails used	Main author	No of People
Gordon	2,975,019	Chrysanthou	25,205	Sheng	1,001
Sutter	288,000	Longtchi	1,260	Moody et al	632
Oner	195,000	Zielinska	887	Gallo	500
Williams	125,284	Bollens	100	Luga	382
Lain	117,864	Wang	10	Gratian	369
Braun	68,743	McElwee	Unavailable	Albladi	316
Taib	56,365			Anawar	252
Bayl Smith	50,000			O'Connor	196
Lain	13,662			Oliveira	158
Toth	13,000			Parsons	117
Steves	9,654			Lawson	102
Musava	4,483			Sarno	69
Koddebusch	3,452			Jones	68
Alwanain	3,000			Das	57
Hanus	2,866			Champa	55
Frank	2,302			Ferreira	52
Yang	2,210				
Kumar	1,774				
Pinto	1,025				
Eftimie	940				
George	876				
Smith	646				
Jalali	430				
Scott	360				
Vishwanath	345				
Sarno	300				
Harrison	194				
Bona	191				
Vishwanath	161				
Abroshan	135				
Bakhshi	97				
Hassandoust	Unavailable				

Table 7 - Analysis of main research type carried out

Two theoretical frameworks were particularly prominent across the selected studies. First, Cialdini's (2007) six principles of persuasion, Reciprocation, Consistency, Social proof, Liking, Authority and Scarcity, were used to analyse the content of phishing emails in several studies (e.g., Bona and Paci, 2020; Champa et al. 2025; Lawson et al. 2020; Musava et al. 2019;

Williams et al. 2018; Zielinska et al. 2016). Second, the Big Five personality trait (BFT) model, comprising Extraversion, Agreeableness, Conscientiousness, Openness and Neuroticism was used to analyse individual susceptibility profiles (e.g., Anawar et al., 2019; Eftimie et al., 2022; Gratian et al., 2018; Lawson et al., 2020; Moody et al., 2017). The use of these frameworks is discussed further in the context of each research question (RQ) below.

3.1.7 Quality of the evidence base

Application of the six-criterion quality assessment framework (Section 2.7) to the 54 included studies produced the following distribution: 27 studies (50%) at the High tier (total score ≥ 5.0), 23 (43%) at the Medium tier (3.5 to 4.5), and 4 (7%) at the Low tier (< 3.5). The corpus mean was 4.62 out of 6. Quality scores were used in the analytical synthesis in Section 4 to identify where specific findings rest on stronger or weaker evidence and to support the analytical comparison summary presented as Table 8. The quality tier of each individual study is reported alongside its findings in the synthesis matrix (Table 3).

3.2 Individual Factors (RQ1)

The literature reviewed identifies four broad categories of individual factors associated with phishing susceptibility: personality traits (Section 3.2.1); cognitive and behavioural factors, including the relationship between stated intentions and actual behaviour (Section 3.2.2); persuasion techniques (Section 3.2.3); and demographic characteristics, principally age and gender (Section 3.2.4).

3.2.1 Personality traits

The BFT is the most widely accepted model to describe the structure of personality traits (Abood, 2019). The BFT are defined as Extraversion, Agreeableness, Conscientiousness, Openness and Neuroticism. Several have been identified as potential predictors of phishing susceptibility across multiple studies.

Extraversion is the most consistently evidenced trait in this context. High extraversion is associated with higher susceptibility across several studies. Although extraversion correlates with some positive behaviours, findings vary by context, and some studies report extraversion, agreeableness and neuroticism characterise non-responders.

Eftimie et al. (2022) also argue that agreeableness, the tendency to avoid conflict and seek cooperation, is the trait most consistently associated with phishing susceptibility because individuals high in this trait may find it harder to reject a request appearing to come from an authority or colleague.

Whilst not one of the BFT, altruism, which is related to but distinct from agreeableness, has also been identified as a vulnerability as phishing emails using altruistic appeals demonstrate a strong correlation with unsafe behaviour because appeals to empathy or helpfulness often bypass recipients' critical scrutiny (Toth et al., 2025).

Conversely, Moody et al. (2017) found few instances where personality traits influenced susceptibility, which they noted was contrary to their own predictions. Together, the evidence for personality-trait-based susceptibility is consistent for extraversion and agreeableness but not universal, suggesting these function as risk factors rather than deterministic predictors.

3.2.2 Cognitive and behavioural factors

Two cognitive factors, impulsivity and self-efficacy, were also cited in the literature reviewed alongside the noteworthy tension of the intention-behaviour gap i.e., the gap between intended behaviour and actual behaviour. As introduced in Section 1, the intention-behaviour gap refers to the disconnect between stated intentions and observed behaviour: people who report intending to behave in a particular way often do not. In the phishing context, employees who report security awareness and express secure intentions during training continue to click on phishing emails under operational conditions.

Impulsivity is a consistent predictor of susceptibility, particularly under time pressure. Self-efficacy shows mixed effects, with some studies reporting protective associations and others finding none.

Impulsivity is a consistent predictor of susceptibility, particularly under time pressure. Self-efficacy shows mixed effects, with some studies reporting protective associations and others finding none. The intention-behaviour gap is clear: users who report high awareness still click on phishing emails.

Vishwanath et al. (2011) identified self-efficacy as a candidate but found it did not predict susceptibility, arguing instead that domain-specific security knowledge was more significant. In contrast, Albladi and Weir (2020) found a combined competence measure (security awareness, privacy awareness and self-efficacy) significantly predicted individual ability to detect social engineering attacks. Vishwanath et al. (2018) argue the mode of cognitive processing is predicated on cyber-risk beliefs and individuals consequently fail to notice cues that could reveal the deception.

A particularly noteworthy behavioural finding is the gap between intended and actual behaviour. Research by Jalali et al. (2019), which they described as one of the first studies to investigate ‘why’ employees click using actual click data, found intended and actual compliance were not linked. This is consistent with findings by Oliveira et al. (2017), who observed a significant discrepancy (especially among older users) between self-reported and actual susceptibility to spear phishing. Hassandoust et al. (2020) and Toth et al. (2025) similarly found large discrepancies between students perceived and actual reactions to phishing messages and George et al. (2020) conclude intentions are often poor predictors of behaviour. This intention-behaviour gap has significant implications for the design of security awareness programmes and is discussed further in Section 4.

3.2.3 Email design and persuasion techniques

The persuasion techniques used in phishing emails consistently influence susceptibility, with urgency and authority the most frequently cited effective cues (Steves et al., 2020). Bona and Paci (2020) found urgency-based emails made employees more vulnerable than those using authority or no psychological principle. Bollens (2024) and Vishwanath et al. (2011) found individuals focus disproportionately on urgency cues, often neglecting the sender address and grammatical quality of an email, a pattern that increases susceptibility. Authority was identified by Koddebusch (2022) as the most effective persuasion technique, followed by urgency, benefit and danger. Wang and Lutchkus (2023) similarly found authority and reciprocation were the most common tactics. Williams et al. (2018) found urgency and authority cues specifically increased susceptibility in workplace settings. In their research, Smith and Dupuis (2025) found

even minimally stylised lures with coherent narratives e.g., job opportunity can elicit engagement when delivered.

However, the relative importance of each persuasion technique varied. Social proof and scarcity cues produced more clicking responses in research by Bayl-Smith et al. (2021), whilst Ferreira and Lanzini (2015) found liking, similarity, and deception most common, with authority ranking only third for data theft. Oliveira et al. (2017) found age moderated effectiveness, younger users being most susceptible to scarcity and older adults to reciprocation. The variation across studies likely reflects differences in sector context, study design and specific cues used, indicating the need for research that can isolate the effects of individual persuasion techniques.

3.2.4 Demographic factors: age and gender

Age and gender are among the most frequently studied demographic predictors of phishing susceptibility. However, the evidence is divided and no consensus has emerged. Regarding gender, six studies found no statistically significant effect (Bona and Paci, 2020; Eftimie et al., 2022; Moody et al., 2017; Koddebusch, 2022; Kumar et al., 2023; Parsons et al., 2013), while seven found women more susceptible (Abroshan et al., 2021; Frank et al., 2022; Hassandoust et al., 2020; Oliveira et al., 2017; Oner et al., 2025; Sheng et al., 2010; Yang et al., 2022). Alwanain (2019) contradicts this, finding most victims to be male. Taib et al. (2019) suggest gender may influence reporting behaviour rather than susceptibility itself, while Luga et al. (2016) point to years of PC use as a potentially contributing variable.

Findings on age are similarly divided. Das et al. (2020) reported older participants less susceptible to spear phishing, O'Connor et al. (2021) found that older adults were more likely to be targets but not victims of fraud compared to younger adults. Conversely, Frank et al. (2022) and Taib et al. (2019) found older employees (particularly those over 41) more vulnerable. Oliveira et al. (2017) found older adults more susceptible to reciprocation-based techniques. The divided evidence on demographic factors most likely reflects diversity across study contexts and indicates demographic characteristics alone are insufficient to profile phishing risk, particularly given the dominance of US and academic settings in existing literature.

3.3 Environmental factors (RQ2)

Contextual factors beyond the individual also contribute to susceptibility, including workload, timing of email delivery and home versus workplace context. A summary is shown in Table 3. Employees with high workloads were more likely to click on phishing (Jalali et al., 2019) and high email load combined with media use significantly increases susceptibility (Vishwanath et al., 2011). Time pressure is a significant factor and reduces attention to CSAT content (Lain et al., 2022) and being busy negatively impacts phishing susceptibility (Williams et al. 2018).

Emails appearing to originate within an organisation were associated with elevated phishing success (Toth et al., 2025), a finding developed further in Section 4. Williams et al. (2018) found online risk behaviour differs between work and home contexts, particularly regarding perceived personal impact and IT support availability, though most factors influencing email trust applied in both settings, with the work environment perceived as more secure.

The time at which phishing emails are delivered has been proposed as an environmental variable, though direct empirical evidence is limited. Eftimie et al. (2022) deliberately avoided morning delivery in their simulation because employee alertness was highest at that time, while Bakhshi (2017) conducted one simulation starting at 15:30 because employees would not have been fully attentive to social engineering indicators. These reflect practitioner assumptions yet to be systematically tested.

The theoretical basis is plausible, attentional resources deplete over the course of a working day (Kahneman, 2011). The effect requires dedicated empirical investigation and is discussed in Section 4.2.

3.4 The effectiveness of organisational controls (RQ3)

The findings indicate opinion also appears to be divided as to the effectiveness and impact that organisational controls and CSAT have on phishing susceptibility.

3.4.1 The benefits of organisational controls

Several studies provided evidence that well-designed organisational controls can meaningfully reduce phishing susceptibility. A study of 1,300 employees across 20 organisations found regular phishing training with immediate feedback improved awareness, reduced impulsive responses, preventing 70% of initially deceived employees from repeating the behaviour (Toth et al., 2025).

Organisations with permanent IT staff and frequent CSAT activities demonstrated better resistance to phishing than those without (Pinto et al., 2022) and research in a student population found a significant positive relationship between phishing awareness and protective practices (Hassandoust et al., 2020). The amount of time employees spent on CSAT was a significant factor in susceptibility, those undertaking more CSAT performing better than those with none or limited CSAT (Frank et al., 2022).

Strategies to increase threat awareness have also been shown to increase the likelihood of protective behaviour, including reporting suspected phishing emails (Bayl-Smith et al., 2021). Employees with lower-income, working in more clerical, monotonous roles were found to be more likely to click on phishing links, suggesting these groups may require targeted CSAT (Hanus et al., 2022). One study observed a decline in the success rate of phishing simulations following CSAT, indicating an overall improvement in organisational awareness (Eftimie et al., 2022). CSAT tailored to be more in line with (what the author described as) the users' circle of relevance was associated with greater improvement over time (Sutter et al., 2022) and prior exposure to phishing education has been associated with reduced susceptibility more broadly (Sheng et al., 2010).

3.4.2 The ineffectiveness of organisational controls

Other studies found limited or no lasting effectiveness from CSAT. One found CSAT only affected behaviours associated with agreeableness and had limited broader impact (Eftimie et al., 2022). Research by Lain et al. (2022) found the content component of embedded phishing

training was not responsible for significant improvements in employee vigilance. Research by Williams et al. (2018) highlighted the complex nature of workplace phishing susceptibility and argued a 'one-size-fits-all' approach to phishing simulations is unlikely to be sufficient.

A second strand of research found little measurable effect from generic CSAT programmes overall. Luga et al. (2016) found little difference between participants who had taken anti-phishing training (67.10%) and those who had not (65.55%) in their ability to distinguish legitimate from phishing pages, though the authors cautioned against drawing firm conclusions given their small sample size. Scott et al. (2024) concluded phishing training methods have little effect overall on end-user negative response actions and that annual CSAT is not effective in mitigating phishing threats. This was similar for Parsons et al. (2013), who argued generic, 'one-size-fits-all' CSAT programmes tend to fail because the needs of individual staff members are not considered.

A third more fundamental concern is that even where CSAT changes user attention, it does not necessarily change user judgement. Harrison et al. (2016) found those who more closely attend to a phishing e-mail do not necessarily end up making the right choice and phishing education simply increased people's suspicion. Phishing simulations can be less effective and sometimes even counterproductive; organisations and researchers should therefore enhance technical detection measures to reduce the number of phishing emails reaching users' inboxes (Braun et al., 2025). These findings are consistent with the intention-behaviour gap identified in Section 3.2.2, suggesting that CSAT approaches relying on knowledge transfer alone are unlikely to produce durable behavioural change.

Figure 3 provides a consolidated overview of the evidence base examined across research questions, framing the discussion that follows in Section 4.

While the preceding analysis highlights discrete variables affecting phishing susceptibility, these factors do not operate in isolation. To synthesize these disparate findings across individual, environmental and organizational domains, we construct a unified conceptual framework. This framework (Figure 4) serves as a structural lens to reconcile conflicting empirical evidence and map the contextual dynamics of user vulnerability.

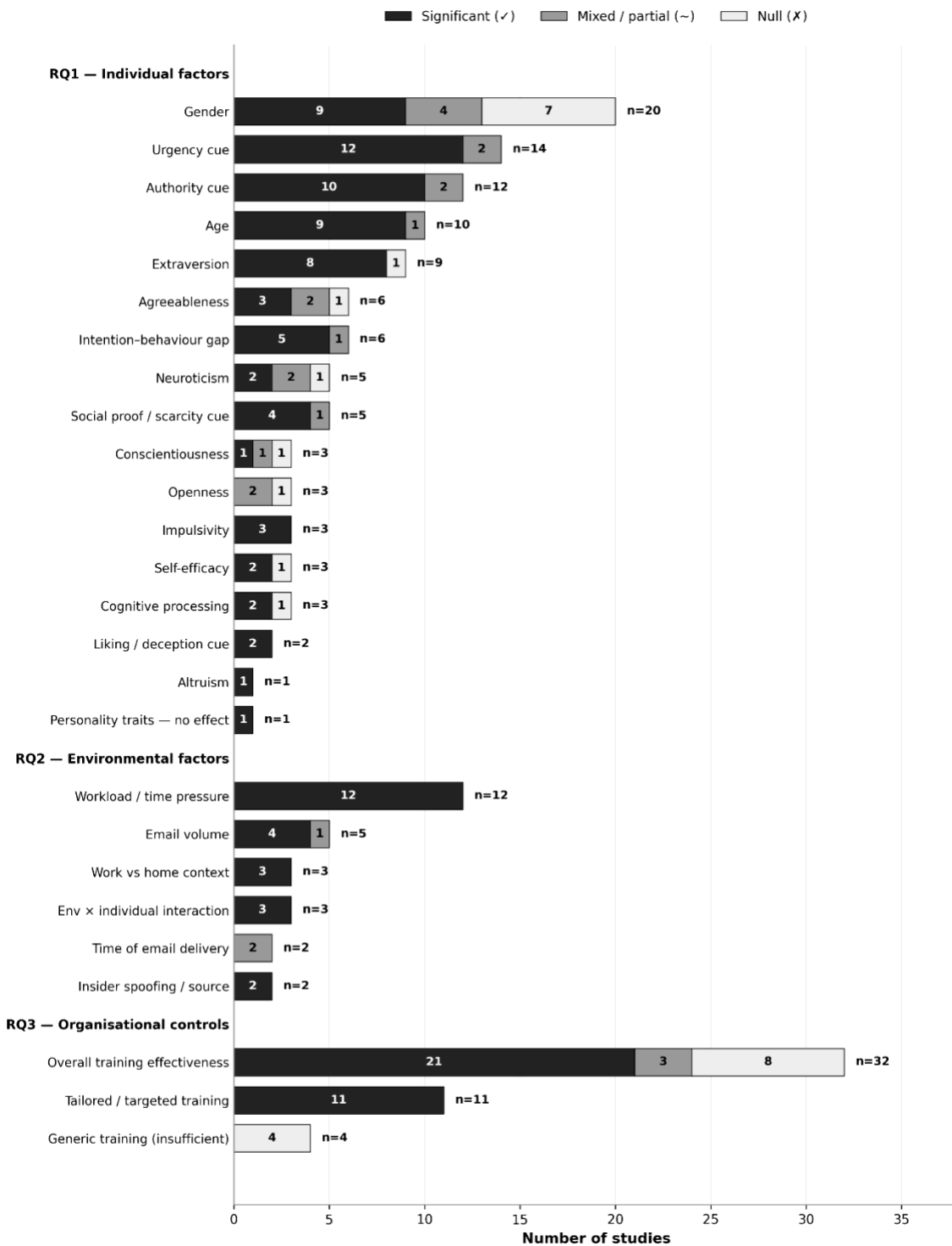


Figure 3 - Number of studies reporting each factor across the 54 reviewed papers

4 Discussion

This section synthesises the findings of the review through three connected analytical moves. First, it identifies the patterns that emerge across the 54 included studies and the verdicts those patterns support on each individual, environmental and organisational factor (Table 8).

Second, it develops a conceptual framework (Figure 4) that integrates these factors into a single model centred on the intention-behaviour gap, which the synthesis identifies as the central failure mode in current understanding of phishing susceptibility. Third, it draws out the practical, theoretical and ethical implications that follow from comparing the structure of susceptibility revealed by the framework with the structure of current organisational defences.

The conceptual framework presented in Figure 4 is developed as a synthesis lens rather than a tested theoretical model. Its function is to organise the empirical patterns identified across the corpus into a coherent structure that supports analysis, rather than to claim a validated causal account. The framework draws together three categories of contributor to phishing susceptibility that prior reviews have typically examined in isolation: individual predispositions (personality, cognitive style, demographics, knowledge), environmental conditions (cognitive load, source context, temporal context, work setting), and the intention-behaviour gap (the disconnect between stated security intentions and observed behaviour, grounded theoretically in habit and automaticity and in dual-process accounts of cognition). These three contributors operate in parallel and interact with each other in shaping the decision process that follows from a phishing stimulus. The framework also makes visible the layer of organisational controls that surrounds this decision process and identifies where current intervention effort concentrates within that layer. The discussion that follows works through each component of the framework in turn, beginning in Section 4.1 with the intention-behaviour gap as the central failure mode.

To explain how the framework works, consider an employee who has completed annual CSAT, expresses confidence in their ability to identify phishing emails but clicks on a phishing link on a busy Friday afternoon. Each of the three contributors in the framework is engaged in this single moment. The employee's individual factors are present but not protective: training delivered weeks earlier formed an intention to be vigilant but not a specific implementation intention for this email at this moment. The environmental conditions of workload, time pressure and high volume of email amplify reliance on System 1 processing. Further the intention-behaviour gap operates independently: routine email handling has become habitual, each email triggering a near-automatic process-and-move-on response that does not engage deliberate intention. The click is not a failure of knowledge, motivation or character, it is the result of the three contributors operating jointly under conditions for which the protective measures were not designed.

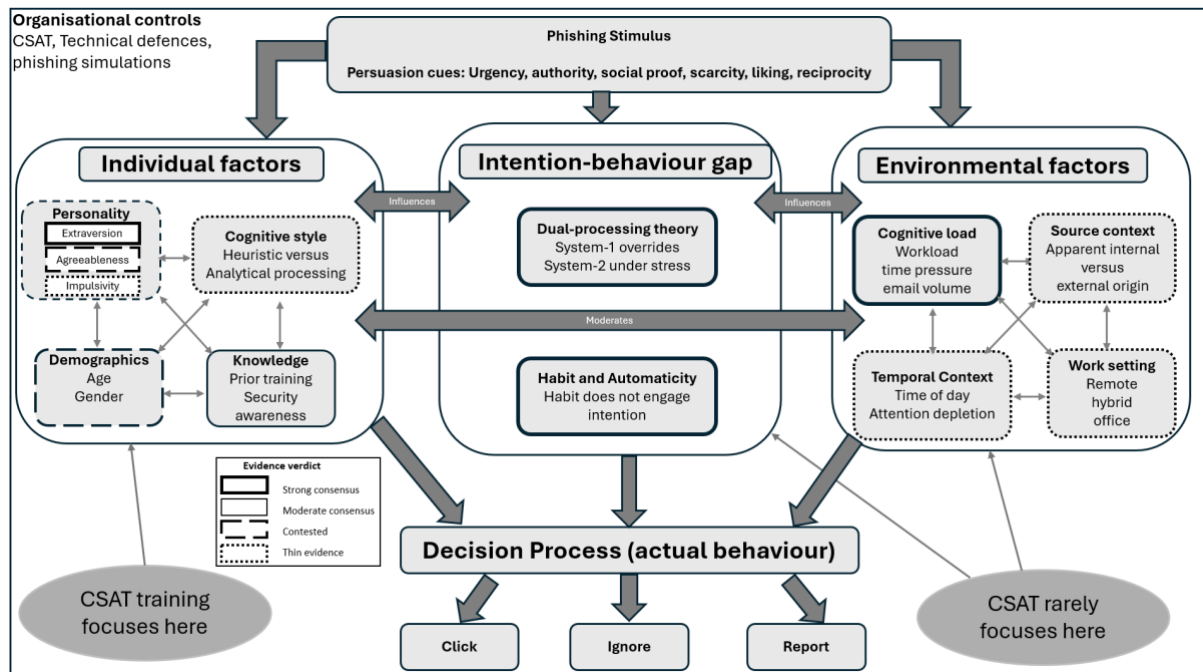


Figure 4 - Conceptual Framework.

The phishing stimulus and its persuasion cues (top) are the external trigger to the decision process; the three contributor boxes (individual factors, intention-behaviour gap, environmental factors) operate as parallel influences within the susceptibility structure. The intention-behaviour gap is present as a contributor because it operates partly independently of individual or environmental conditions, via habit and automaticity (see Section 4.1). Sub-elements within each contributor box are illustrative examples drawn from the reviewed literature, not a definitive enumeration. Bold arrows denote causal flow (from the stimulus through the contributors to the decision process and behaviour). Thinner bidirectional arrows denote interaction relationships between contributors. Sub-element border styles indicate evidence verdict from the synthesis, solid = consensus support, dashed = contested, dotted = thin evidence.

Table 8 presents an analytical summary of the major findings against their vote-counting evidence base and mean quality assessment score, anchoring the discussion that follows. This is an aggregated verdict per factor, drawn from the study-level matrix in Table 3, with mean methodological quality and analytical verdict added.

Factor	n	✓	X	~	Mean QA Score	Evidence pattern
RQ1 — Individual factors						
Extraversion	9	8	1	0	4.78	Predominantly significant
Agreeableness	6	3	1	2	4.75	Mixed / contested
Conscientiousness	3	1	1	1	4.5	Inconclusive (small n)
Neuroticism	5	2	1	2	4.9	Inconclusive
Openness	3	0	1	2	4.5	No clear effect
Altruism	1	1	0	0	6	Single study
Personality traits — no effect	1	1	0	0	5	Single study
Impulsivity	3	3	0	0	3.5	Consistent (small n)
Self-efficacy	3	2	1	0	4.83	Inconclusive (small n)
Cognitive processing	3	2	1	0	5.33	Inconclusive (small n)
Intention–behaviour gap	6	5	0	1	5.17	Strong, consistent
Urgency cue	14	12	0	2	4.25	Strong, consistent
Authority cue	12	10	0	2	4.33	Strong, consistent
Social proof / scarcity cue	5	4	0	1	4.5	Strong, consistent
Liking / deception cue	2	2	0	0	3.25	Consistent (small n)
Gender	20	9	7	4	4.62	Mixed / contested
Age	10	9	0	1	4.85	Significant; direction unresolved
RQ2 — Environmental factors						
Workload / time pressure	12	12	0	0	4.79	Strong, consistent
Email volume	5	4	0	1	4.4	Consistent
Time of email delivery	2	0	0	2	4.5	Thin; no clear effect
Insider spoofing / source	2	2	0	0	6	Consistent (small n)
Work vs home context	3	3	0	0	4.83	Consistent (small n)
Env × individual interaction	3	3	0	0	5.33	Consistent (small n)
RQ3 — Organisational controls						
Overall control / training effectiveness	32	21	8	3	4.8	Predominantly effective; contested
Tailored / targeted training	11	11	0	0	4.95	Strong, consistent
Generic training (insufficient)	4	0	4	0	4.62	Consistently ineffective

Table 8 - Analytical summary of main findings. Each factor identified across the 54 included studies is shown with its vote-count breakdown (significant / null / mixed), mean methodological quality score (from the Section 2.7 quality assessment), and the resulting evidence verdict. Notes: Vote-count shows the number of studies reporting significant (✓), null (X), or mixed/partial (~) findings for each factor. Mean quality is the average six-criterion quality assessment score (Section 2.7) of studies addressing each factor.

4.1 The intention-behaviour gap as the central failure mode

In response to RQ1, the evidence indicates that several individual factors are associated with increased phishing susceptibility, with extraversion, agreeableness and impulsivity emerging consistently across multiple studies (Anawar et al., 2019; Eftimie et al., 2022; Gratian et al., 2018; Lawson et al., 2020; Longtchi et al., 2024). However, the most analytically consequential finding from the review is not the list of individual factors but the observation across multiple

studies that stated security intentions do not reliably predict actual clicking behaviour (George et al., 2020; Jalali et al., 2019; Oliveira et al., 2017; Toth et al., 2025). Employees who self-report awareness of phishing risks and express secure intentions do not translate these into protective behaviour under operational conditions. This gap between intention and behaviour is a central failure mode the review identifies in the existing evidence base. Critically, the gap operates as a contributor to susceptibility in its own right alongside individual and environmental factors rather than as a downstream consequence of them. A person without measurable individual risk factors and without environmental amplification can still fall through the gap when confronted with a phishing stimulus because the cognitive mechanisms underpinning the gap are partly independent of context. Marin et al. (2023) suggest that high-conscientiousness individuals may consider reporting outside of their duties or misunderstand it, illustrating that even motivated, security-aware employees are not immune.

The intention-behaviour gap has well-developed theoretical foundations in behavioural science that have not been substantively engaged with in the phishing-susceptibility literature. The most directly applicable mechanism is habit and automaticity (Wood and Neal, 2007). Routine email handling becomes habitual, with most emails processed without deliberate engagement of intentional decision-making. Under this view, intention is not bypassed by failure of will but by the simple fact that habitual behaviour does not engage intention at all. This mechanism operates independently of individual differences and environmental conditions, a vigilant, well-rested employee handling email in a low-pressure setting can still click on a phishing message if email-clicking has become a routine, automatic response to inbox stimuli.

The dual-process framework (Kahneman, 2011) provides a complementary account. Stated intentions are typically formed under System 2 conditions (reflective, deliberate cognition), often during awareness training when security is an explicit focus. Real-time phishing decisions are typically made under System 1 conditions (fast, heuristic, automatic). Environmental conditions such as time pressure, cognitive load or distraction, further amplify System 1 dominance but the gap exists at baseline even when these amplifiers are absent. A further relevant mechanism, implementation intentions (Gollwitzer, 1999), distinguishes general intentions from specific situational plans such as unexpected emails asking to click a link. Generic awareness training forms the former; only practice-based, scenario-specific training forms the latter, a distinction that helps explain the training-effectiveness pattern discussed in Section 4.3.

These theoretical foundations have substantive implications for organisational practice. Training that forms general intentions is unlikely to change behaviour in the operational moments where phishing decisions occur. Interventions that form implementation intentions (specific if-then plans practised under realistic conditions) and that interrupt habitual processing (friction at the decision point) are more likely to be effective. The conceptual framework (Figure 4) positions the intention-behaviour gap as a contributor to the decision process that is influenced by, and in turn influences, individual factors and environmental conditions.

4.2 Why the demographic and individual-factor evidence is contested

Several RQ1 factors show divided evidence: gender (9 of 20 studies finding effects, 7 null, 4 mixed), age direction (9 of 10 finding effects with conflicting directions), and persuasion-cue prominence. The pattern of divergence reveals different sources for each.

As reported in Section 3.2.4, the evidence on gender is divided. Ribeiro et al. (2024) argue gender is not a meaningful predictor of susceptibility and that other factors should be prioritised in research and intervention design. The present review's evidence is consistent with this position. Studies finding significant gender effects include real-world simulations (Abroshan et al., 2021; Frank et al., 2022; Hassandoust et al., 2020; Oner et al., 2025), lab experiments (Oliveira et al., 2017), and survey-based studies (Sheng et al., 2010). Studies finding no gender effect similarly include real-world simulations (Bona and Paci, 2020; Koddebusch, 2022; Kumar et al., 2023; Eftimie et al., 2022) and lab or survey designs (Moody et al., 2017; Parsons et al., 2013). Both groups contain studies across university and workplace populations. No clear methodological pattern distinguishes the two groups, and the mean quality assessment score of studies addressing gender (4.62, equal to the corpus mean) indicates the divergence is not driven by quality differences either. This suggests that gender effects, where present, are weak and inconsistent, likely confounded by factors not consistently controlled for in the included studies such as role, seniority, and technological familiarity.

The age evidence shows a different pattern. Of the ten studies addressing age, nine find significant effects and one finds mixed effects; no studies find null effects. The divergence is therefore not whether age matters but in what direction. Oliveira et al. (2017) document a non-linear relationship in which younger and older users are vulnerable for different reasons (younger users more susceptible to social-proof cues; older users more susceptible to authority cues). Apparent contradictions in direction-of-effect between studies may reflect which element of the age distribution each study sampled rather than genuine disagreement about whether age matters.

Email cues such as urgency, authority, and social proof are well evidenced persuasion techniques exploited in phishing emails. Many studies confirm their effectiveness (Bakhshi, 2017; Bollens, 2024; Bona and Paci, 2020; Champa et al., 2025; Steves et al., 2020; Williams et al., 2018). The corpus shows clear consensus on whether these cues work: 12 of 14 studies addressing urgency report significant effects, 10 of 12 addressing authority do likewise, with no studies finding cues ineffective at the corpus level. Where the literature divides is on which cue is most powerful but closer inspection suggests this is partly a function of what each study measured. Koddebusch (2022), measuring real-world click rates in a workplace simulation, identifies authority as the most effective cue at producing susceptibility. Ferreira and Lanzini (2015), analysing the cues used in real phishing emails, found liking and deception to be the most prevalent in attacker behaviour. These findings answer different questions (target susceptibility vs. attacker behaviour) and are not in direct contradiction. The consistent finding that urgency and authority cues cause individuals to overlook other email elements, such as sender address and grammar (Vishwanath et al., 2011), has direct relevance to how CSAT is designed, implying CSAT should specifically focus on habits of deliberate scrutiny even under perceived time pressure.

Three implications follow. The contested evidence on gender reflects a genuinely weak or confounded effect rather than a methodological artefact resolvable by future studies. Age direction is consistent with Oliveira et al.'s (2017) non-linear effect and points to the importance of sampling the full adult age range. Demographic profiling is therefore not an evidence-based foundation for security control design, a practical implication developed in Section 4.5.

4.3 Environmental amplifiers and the training mismatch

In response to RQ2, the findings from the reviewed studies indicate that environmental factors, principally workload, time pressure, email volume, the timing and apparent source of emails

and the work versus home context exert a meaningful influence on phishing susceptibility. However, this area remains considerably less researched than individual psychological and demographic factors. Critically, evidence suggests environmental pressures do not operate in isolation but interact with individual characteristics to either amplify or, in some conditions, attenuate susceptibility a moderation relationship made explicit in the conceptual framework.

4.3.1 Workload, time pressure and cognitive amplification

The most consistently evidenced environmental factor is high workload and its associated time pressure (Jalali et al., 2019; Lain et al., 2022; Sarno et al., 2022; Vishwanath et al., 2011). Jalali et al. (2019), using actual employee click-data, found high-workload employees significantly more likely to click on phishing emails, notable because the same study also showed stated and actual behaviour were not linked. Vishwanath et al. (2011) extended this by showing that high email volume, especially combined with simultaneous media use, substantially increased susceptibility, attributing the effect to cognitive depletion. Williams et al. (2018), Jones (2016) and Sarno et al. (2020) report similar effects, while Gallo et al. (2024) found users who spend more time reading emails are less susceptible and propose that disabling link clicking briefly after opening could be an effective preventive technique.

These findings point to a common underlying mechanism: high workload reduces attentional resources available for deliberate scrutiny, pushing decision-making towards faster, less reflective processing. This is consistent with the dual-process account of decision making with Kahneman's (2011) distinction between fast, intuitive System-1 responses and slower, analytical System-2 processing with time pressure systematically favouring the former, the same theoretical mechanism developed in Section 4.1 in relation to the intention-behaviour gap. Lain et al. (2022) extend this picture by demonstrating time pressure also reduces engagement with CSAT content, creating a compounding effect. The employees most exposed to phishing risk are also the least receptive to the training designed to protect them. This suggests short, contextually relevant micro-training delivered at lower-workload moments may be more effective than longer programmes delivered at fixed intervals.

4.3.2 Email volume, source context and working environment

Related to workload is the question of email volume and the attentional strategies employees adopt when processing large numbers of emails. Bollens (2024) and Vishwanath et al. (2011) both found individuals confronted with high email volumes tend to focus disproportionately on urgency and surface-level cues whilst neglecting more reliable indicators of legitimacy such as sender domain, grammatical quality and the plausibility of the request. Steves et al. (2020) similarly identified urgency as a key predictor of phishing success and Bona and Paci (2020) found urgency-based emails made employees more vulnerable than those using other psychological principles. Users with intermediate job experience (3 to 11 years) appear more resistant to urgency-based attacks than those with less or more experience, possibly because highly experienced employees hold greater responsibility and face more genuinely urgent messages (Gallo et al., 2024).

The combination of high email volume and urgency-based phishing techniques is particularly potent: the volume creates the attentional conditions under which urgency cues are most effective. Phishing simulations should therefore systematically vary the volume and timing of simulated emails to reflect realistic workload conditions.

A second contextual amplifier is the apparent source of emails. Toth et al. (2025) found internally spoofed emails were disproportionately effective because they exploit the trust employees place in communications from colleagues or management. Williams et al. (2018) similarly found employees perceived the work environment as more secure than the home

environment, which may create a false sense of security that increases susceptibility to internally spoofed communications. This attack method is particularly difficult to counter through awareness training alone because it exploits an entirely rational heuristic, trusting internal communications that ordinarily serves employees well.

A third contextual factor is the work-vs-home environment. Williams et al. (2018) found that whilst online risk behaviour can differ between work and home, particularly regarding perceived personal impact and IT support availability, most factors influencing email trust applied equally in both. The home context does not straightforwardly reduce vigilance but the absence of institutional security infrastructure may leave home-based workers less equipped to act on recognised threats. Few of the 54 reviewed studies examined remote or hybrid working specifically, despite the shift in working patterns since 2020. The environmental pressures of remote work, less structured hours, greater task-switching, reduced colleague consultation, blurred work-life boundaries, likely to differ from those in earlier office-based studies and represent a significant gap in the current evidence base.

4.3.3 The interaction between environmental and individual factors

A theme running across the environmental findings is that contextual pressures do not operate independently of individual characteristics but interact with them. A highly impulsive employee may be significantly more susceptible under high-workload conditions than under low-workload conditions, whereas a more conscientious individual may maintain scrutiny even under pressure. Conversely, a generally vigilant employee may nonetheless be caught by internally spoofed emails that exploit contextually legitimate trust. Jalali et al. (2019) and Toth et al. (2025) both found large discrepancies between employees stated intentions and actual behaviour and these were most pronounced under conditions of cognitive and situational pressure, suggesting that stated intentions are simply less reliable predictors of behaviour when environmental demands are high. This interaction between individual vulnerability and environmental conditions means research designs examining either category of factor in isolation are likely to produce an incomplete and potentially misleading picture of real-world susceptibility. This is shown in Figure 4 explicit by the conceptual framework's moderation relationship between the individual factors and environmental factors layers.

4.3.4 The training mismatch

In response to RQ3, the evidence on the effectiveness of organisational controls, principally CSAT and phishing simulations, is also divided. Whilst several studies demonstrate measurable improvements in employee vigilance following CSAT and simulation programmes (Eftimie et al., 2022; Frank et al., 2022; Hassandoust et al., 2020; Pinto et al., 2022; Sheng et al., 2010; Sutter et al., 2022; Toth et al., 2025), others find little or no lasting effect (Bona and Paci, 2020; Lain et al., 2021; Luga et al., 2016; Parsons et al., 2013; Scott et al., 2024). As with the demographic evidence discussed in Section 4.2, the variance reflects substantive differences in what is being evaluated.

Studies finding training effective share three characteristics: the training is tailored to role and prior experience (Sutter et al., 2022; Toth et al., 2025; Frank et al., 2022; Musava et al., 2019); it is frequent and embedded, with feedback at the point of simulated failure (Toth et al., 2025; McElwee et al., 2018; Pinto et al., 2022) and it forms implementation intentions rather than general intentions through scenario-based practice (Bayl-Smith et al., 2021; Sheng et al., 2010). The vote-counting evidence is unambiguous: 11 of 11 studies of tailored, scenario-based training find it effective (mean quality 5.0/6).

Toth et al. (2025) in a study with embedded feedback, prevented 70% of initially deceived employees from repeating the behaviour. Sutter et al. (2022), with 288,000 simulated emails,

found role-tailored training to be associated with greater improvement than generic programmes. Frank et al. (2022) found volume of CSAT was a significant predictor of performance, and Pinto et al. (2022) found organisations with permanent IT staff and frequent training showed better overall phishing resistance.

Studies finding training ineffective evaluate a markedly different design. Scott et al. (2024) concluded that phishing training methods have little overall effect on end-user behaviour and specifically identified annual CSAT as ineffective. This is consistent with Parsons et al. (2013), who argued that generic, one-size-fits-all programmes fail because they do not account for the differing needs of individual staff. Luga et al. (2016) similarly found no meaningful difference between trained and untrained participants. Most revealingly, Lain et al. (2022) found that the content component of embedded phishing training was not responsible for significant improvements in vigilance, implying that the act of completing training may account for whatever short-term improvement is observed. Eftimie et al. (2022) found that training only affected behaviours associated with agreeableness, suggesting a single programme cannot address the full range of individual vulnerabilities. Vote-counting for generic annual training is also unambiguous: 0 of 4 studies find this effective (mean quality 4.5/6) a robust null finding.

Taken together, these findings reframe the debate: the question is not whether training works in principle but whether currently delivered annual, generic, knowledge-focused training is fit for purpose given what is known about the psychological, cognitive and environmental roots of susceptibility (Williams et al., 2018).

This pattern fits the theoretical account developed in Section 4.1. Generic training forms general intentions; tailored, scenario-based training forms implementation intentions. Generic training engages System 2 (in the training session) but does not interrupt System 1 (in the operational moment), scenario-based practice begins to build the automaticity of the secure response itself. Generic training is also delivered infrequently and at fixed intervals, typically annually, making it particularly weak against the habit-and-automaticity mechanism Wood and Neal (2007) identify, which requires repeated practice in the cued context to disrupt established routines.

Few of the organisational controls examined in the reviewed studies appear to have been explicitly designed with the psychological, demographic, or environmental factors identified under RQs 1 and 2 in mind. If impulsivity and agreeableness are among the strongest individual predictors (Longtchi et al., 2024; Eftimie et al., 2022; Jones, 2016; Parsons et al., 2013), training focused on factual recognition addresses the wrong target. An impulsive employee may know what a phishing email looks like but still click under time pressure because the failure mode is attentional rather than informational. Training that builds pause-and-check habits under simulated cognitive load would be better aligned. Hanus et al. (2022) similarly found employees in lower-income, more clerical and monotonous roles are more likely to click on phishing links, a group likely to be under-represented in the design assumptions of standard CSAT. Directing training resources proportionately towards higher-risk employee groups, rather than distributing them uniformly, would be a straightforward and currently underutilised application of the available evidence.

The conceptual framework (Figure 4) makes the mismatch visually explicit: current organisational controls predominantly target the individual factors layer while susceptibility is driven jointly by individual predispositions, environmental amplifiers and the intention-behaviour gap at the decision point.

4.4 The limits of the current evidence base

A cross-cutting finding is the limited real-world applicability of the existing evidence base along three dimensions: setting, geography and industry.

Setting. Twenty-five of the 54 studies were conducted in universities or schools, where student populations differ from employees in age, technological familiarity, organisational stakes and research participation incentives. The assumption that findings transfer directly to industry settings is therefore questionable, as several authors acknowledge. Where large-scale simulations have been conducted, Gordon et al. (2019) with nearly 3 million emails, Sutter et al. (2022) with 288,000, Williams et al. (2018) with 125,284, the focus has been quantitative, with limited researcher control over the content and persuasion techniques used. Other researchers (Bollens, 2024; Chrysanthou et al., 2024; Longtchi et al., 2024; Wang and Lutchkus, 2023) have relied on third-party library datasets with similar control limitations.

Geography. Research is also heavily concentrated in WEIRD societies, with 21 of the 54 studies based in the United State alone. Cultural factors that may influence susceptibility, attitudes towards authority, digital literacy, organisational culture, vary considerably across national contexts. Hasegawa et al., 2024 similarly identify WEIRD dominance as a constraint on generalisability in adjacent privacy and security research. The implications for non-WEIRD contexts admit specific theoretical predictions: authority-cue effects may be systematically larger in high-power-distance cultures than in the WEIRD samples that dominate the existing evidence base (Hofstede et al., 2010); gender-effect findings established in relatively gender-egalitarian workplace contexts may not transfer to cultures with different gender and technology norms; social-proof effects may operate differently in collectivist than individualist cultures. Cross-cultural empirical research is identified as an explicit priority in Section 4.6.

Industry. The corpus is heavily weighted toward sectors with relatively high cybersecurity maturity (financial services, healthcare, large IT-related companies) and toward educational settings. The implications of the findings for manufacturing, agriculture, retail, hospitality, construction, public-sector and SME contexts are not directly evidenced. In particular, the training-effectiveness literature is almost entirely based on organisations with substantial training infrastructure already in place; what works for an enterprise with a dedicated security team and routine simulation programme may not transfer to a small business with no security personnel and no prior training.

The cumulative effect of these three concentrations is that the existing evidence base predicts phishing susceptibility most reliably for educated workers in WEIRD societies employed by high-maturity organisations in finance, healthcare or technology. For other contexts, the findings should be treated as mechanisms to be tested rather than as direct prescriptions. The conceptual framework applies in principle across these contexts, but the magnitudes and moderating relationships it specifies are likely to require local empirical calibration before deployment.

4.5 Implications

4.5.1 Theoretical implications

The review makes three theoretical contributions. The first and most consequential is the identification of a structural mismatch between current organisational controls and the multi-factor structure of phishing susceptibility. Cyber security awareness training predominantly targets individual knowledge through information transfer. The review's evidence shows susceptibility is driven not by individual knowledge alone but jointly by individual predispositions, environmental conditions and the intention-behaviour gap. Organisational defence design that targets only the individual layer leaves the other two contributors

unaddressed helping to explain why phishing remains the leading cause of organisational data breaches.

The second contribution is the conceptual framework (Figure 4) that integrates these three contributors. The framework positions the phishing stimulus and its persuasion cues as the trigger, individual factors and environmental factors as parallel contributors to the decision process and the intention-behaviour gap as a third contributor, influenced by and in turn influences the other two. The gap is grounded in two complementary theoretical accounts: habit and automaticity (Wood and Neal, 2007), which explains why routine email handling can bypass intention independently of context and dual-process theory (Kahneman, 2011), which explains how System 1 processing comes to dominate under cognitive load. The framework is presented as a synthesis lens for organising the empirical patterns in the reviewed literature rather than as a tested theoretical model.

The third contribution is the systematic evidence verdict on each factor (Table 8), produced by combining vote-counting (Section 2.8) with the quality assessment of each included study (Section 2.7). The verdict distinguishes factors with strong consensus support (workload, tailored training, the intention-behaviour gap itself) from those that are contested (gender, age direction, persuasion-cue prominence), robust nulls (generic training) or supported by only thin evidence (altruism, insider spoofing). This distinction provides a more nuanced foundation for both research priorities and practical recommendations than the descriptive synthesis present in prior reviews of the field.

4.5.2 Implications for practice

The review supports five practical implications for organisational cybersecurity practice. Each is grounded in specific findings of the review and stated at a level of generality that allows application across industries, with the caveats about under-represented contexts noted in Section 4.4.

1. Replace generic annual CSAT with tailored, scenario-based training. The robust null finding on generic annual training (4 of 4 studies, mean quality 4.62/6) and the strong consensus on tailored, scenario-based training (11 of 11 studies, mean quality 5.0/6) together support a decisive recommendation: organisations should replace one-size-fits-all annual training with role-tailored, scenario-based practice that forms implementation intentions rather than general intentions.

2. Treat workload as a security risk. Workload and time pressure are among the most consistently evidenced amplifiers of susceptibility (12 of 12 studies). Practitioners should treat email-handling workload as a security risk, not just a productivity issue, and design organisational systems and email-throughput expectations accordingly. Just-in-time micro-training delivered at lower-workload moments is likely to be more effective than longer training at fixed intervals.

3. Build friction at the decision point. The intention-behaviour gap implies that user vigilance alone is insufficient: organisations should design technical and procedural friction at high-risk decision points (external-email banners, contextual warnings, link sandboxing, mandatory verification for high-risk requests) rather than relying on individual attention. The conceptual framework identifies the decision process as the locus of the gap and therefore the appropriate target for friction-based intervention.

4. Avoid demographic-based defence design. The contested evidence on gender and age effects (Section 4.2) means that demographic targeting is not an evidence-based foundation for security control design. Organisations should prioritise behaviour-based and context-based defences over demographic risk profiling. Hanus et al.'s (2022) finding on role-based

susceptibility (clerical and monotonous roles being higher-risk) is an evidence-supported exception: directing training resources by role rather than by demographics is defensible.

5. Strengthen reporting infrastructure and culture. Reporting, both the technical mechanism and the cultural conditions that support it emerges from the review as a more achievable behavioural target than complete prevention of clicks. Organisations should invest in low-friction reporting tools and in cultures that reward reporting rather than punishing clicks (Marin et al., 2023; Bayl-Smith et al., 2021).

4.5.3 Implications for AI-enabled phishing

The findings of this review are based on studies conducted predominantly before the widespread availability of generative AI for attack generation. Generative AI has eroded the surface-level detection cues (grammar, phrasing, generic greetings) on which much existing awareness training is built (Jabir et al., 2025). The persuasion cues identified in RQ1 remain operative; what has changed is the polish and contextual appropriateness of the delivery.

Three implications follow. First, the practical recommendation against generic awareness training is strengthened, not weakened, by AI-enabled phishing. Training that teaches people to spot linguistic indicators is substantially weaker than the pre-AI literature suggested. Second, the intention-behaviour gap is likely to widen, even motivated users struggle to detect AI-generated phishing on surface features, increasing the importance of structural defences over individual vigilance. Third, contextual evaluation matters more than ever, with decisions resting less on surface features and more on whether the email was expected, fits normal practice, or creates unusual urgency (Section 4.6.8).

4.6 Future research agenda

The following research agenda is drawn from the gaps and contradictions identified in this review.

4.6.1 Large-scale, controlled, industry-based phishing simulation studies

The most pressing gap is the absence of large-scale phishing simulations conducted in real-world organisational settings with researcher control over content, persuasion techniques and timing (Gordon et al., 2019; Sutter et al., 2022; Williams et al., 2018). Such studies should collect both quantitative behavioural data and qualitative 'why' data and longitudinal designs spanning at least 12 months would enable assessment of how susceptibility changes over time in response to interventions.

Future research questions:

- How does susceptibility vary across industry sectors when researcher controlled phishing cues are deployed across comparable organisations.
- How does susceptibility evolve over a 12-month period within a single organisation and what factors drive observed changes?

4.6.2 Investigation of the intention-behaviour gap in operational settings

Self-reported security intentions that do not reliably predict actual behaviour under real phishing conditions was a consistent finding across the literature (George et al., 2020; Jalali et al., 2019; Oliveira et al., 2017; Toth et al., 2025). What remains poorly understood are the mechanisms underlying this gap, the cognitive, emotional, and situational factors that intervene between an employee's stated intention and their actual behaviour when clicking and the intervention designs that demonstrably reduce the gap.

Future research questions:

- How large is the discrepancy between self-reported security intentions and actual clicking behaviour in operational workplace settings?
- Which individual, situational and organisational factors moderate the size of the intention-behaviour gap?

4.6.3 Tailored and adaptive CSAT

Given the consistent evidence that generic, one-size-fits-all training is insufficient (Parsons et al., 2013; Scott et al., 2024; Williams et al., 2018), research is needed that designs and rigorously evaluates personalised training interventions targeting individual personality and cognitive profiles, delivered at regular intervals with real-time feedback at the point of simulated phishing failure. The long-term impact should be assessed over at least 12 months.

Future research questions:

- Which individual factors (personality traits, cognitive processing styles, intention-behaviour gap) provide the most useful basis for tailoring training and how can these be measured?
- Does training tailored to an individual's personality and cognitive profile produce greater and more durable reductions in clicking behaviour than generic training of equivalent duration?

4.6.4 Environmental and contextual interaction effects

Future research should investigate the interaction between individual and environmental factors, particularly how time pressure, email volume and fatigue moderate the relationship between personality traits on susceptibility. Two related contexts warrant investigation, remote and hybrid working since 2020 and the time at which phishing emails arrive where afternoon cognitive fatigue may reduce attentional resource (Eftimie et al., 2022; Bakhshi, 2017).

Future research questions:

- How does susceptibility change across the working day and under different workload conditions?
- Does phishing susceptibility differ meaningfully between fully remote, hybrid and office-based working patterns?

4.6.5 Research beyond Western populations

The concentration of existing research in Western and US academic settings constrains generalisability, mirroring a documented bias in broader behavioural science (Hasegawa et al., 2024). Future research should replicate key findings in non-Western, non-English-speaking, economically diverse contexts and in Africa, Asia and South America, where phishing threats are increasingly prevalent but research is sparse.

Future research questions:

- Do the personality and demographic predictors of phishing susceptibility identified in Western populations replicate in non-Western contexts?
- How does the relative effectiveness of persuasion cues vary across cultural contexts?

4.6.6 Bridging the gap between research findings and organisational practice

Few reviewed organisational controls appear to have been designed with the susceptibility factors identified in RQs 1 and 2 in mind. Future research should examine how far practitioners apply susceptibility research when designing CSAT and simulations, with researcher-practitioner co-design offering a productive model.

Future research questions:

- How does the effectiveness of CSAT designed through co-design between researchers and practitioners compare with that of off-the-shelf commercial training products?
- To what extent do security practitioners who design CSAT and phishing simulation programmes draw on academic findings about individual susceptibility?

4.6.7 Ethical and socio-technical open questions

The agenda above is predominantly empirical. Several ethical and socio-technical questions also warrant investigation.

Ethical questions. Phishing simulations are typically conducted without participant consent, raising questions about the trade-off between methodological realism and the psychological contract between employer and employee. Responsibility for employee clicks is contested between the individual, the organisation and technical infrastructure that allowed the email through. Click data is sensitive personal data and warrants protection. Targeting stable individual traits such as impulsivity or agreeableness raises ethical questions about intervention design?

Socio-technical questions. The division of labour between technical controls (filtering, sandboxing) and human-factor controls (training, awareness) warrants further investigation as does the socio-technical impact or remote or hybrid working on phishing decisions. The unintended consequences of phishing simulation programmes, particularly whether fear-based training erode the trust required for legitimate internal communications represent a further question.

4.6.8 Phishing susceptibility under AI-enabled attack conditions

Most studies in the corpus were conducted before the widespread availability of generative AI for attack generation. As argued in Section 4.5.3, AI-generated phishing erodes surface-level detection cues while leaving persuasion-cue mechanisms operative. Replication of key findings, under AI-enabled attack conditions, particularly training effectiveness, demographic effects and the intention-behaviour gap is an urgent research need. AI-enabled phishing also poses a substantive test of the framework: whether the moderating relationships between individual and environmental factors, the centrality of the gap, and the organisational-controls mismatch hold in conditions the existing evidence base does not cover.

4.7 Limitations

This SLR is subject to several limitations. First, although PRISMA was followed rigorously and inclusion/exclusion decisions were guided by pre-defined criteria, the review was conducted predominantly by a single reviewer, leaving some subjectivity in the relevance assessment for papers on the boundary of the defined scope. To mitigate single-reviewer bias, inclusion criteria were defined a priori before screening commenced and were not revised retrospectively; conservative inclusion was applied at the title-abstract stage (papers retained for full-text review whenever doubt existed); borderline full-text decisions were discussed with the second author. Further, independent spot checks were made by the second author. The single-reviewer screening approach remains a methodological constraint of the review and a candidate target for future replication studies.

Second, the search was limited to seven databases and to English-language paper. Although the databases are among the most comprehensive in computing, psychology and information

systems disciplines, relevant non-English or non-indexed studies may have been missed. The reverse snowballing process identified 17 additional papers, suggesting the initial database search was not fully exhaustive and one further recent paper, was identified via ongoing Google Scholar alerts.

Third, the review predominantly covers 2016 to 2025, selected based on the publication from 2016 onwards. Influential earlier work (Ferreira and Lanzini, 2015; Parsons et al., 2013; Sheng et al., 2010; Vishwanath et al., 2011) was included only through reverse snowballing consistent with practice in related SLRs (Desolda et al., 2022). Other pre-2016 work may not have been captured.

Fourth, the diversity of included studies in research design, sample characteristics, phishing cues, outcome measures and sector context makes firm quantitative conclusions difficult. The findings are therefore primarily qualitative and interpretive; future reviews focused on more narrowly defined questions may be better placed to produce quantitative syntheses. The vote-counting synthesis used in this review (Section 2.8) cannot statistically weight studies by sample size or quality in the manner that meta-analysis can. The quality assessment introduced in Section 2.7 partially mitigates this by allowing the discussion to identify where findings rest on higher or lower-quality evidence, but the residual limitation remains.

Fifth, the review is subject to publication bias: studies with statistically significant findings are more likely to enter the published record than studies with null or unimpressive findings (Rothstein et al., 2005). The phishing-susceptibility literature is particularly vulnerable to this bias because simulation studies are resource-intensive and often face organisational pressure to report effects, training-effectiveness studies are exposed to vendor-related reporting bias and the methodological heterogeneity of the field makes null findings easier to under-report. Across the corpus, 28 of the reported findings are null (denoted X in Table 3) as well as 22 partial/mixed (denoted $\sim$ in Table 3), suggesting that null or partial results are at least partly representable in the literature; the residual risk that apparent strong findings are inflated by unpublished null results remains and is acknowledged.

5 Conclusion

This systematic review synthesised 54 peer-reviewed studies on human susceptibility to phishing using the PRISMA framework, organised around three research questions covering individual factors, environmental factors and organisational controls. Its central contribution is the integration of three contributors, individual factors, environmental factors, and the intention-behaviour gap into a single conceptual framework that explains why phishing remains the most common cause of organisational data breaches, and that makes visible the structural mismatch between where current controls land and where susceptibility actually lives.

The evidence base, while substantial in volume, is limited in its real-world applicability by its reliance on educational settings, Western populations and high-maturity sectors. Closing the gap between research and practice will require large-scale studies in real-world industry environments, across a broader range of national, cultural and industry contexts.

The persistence of phishing reflects genuine individual vulnerabilities but it is also a failure of organisational imagination about what those vulnerabilities demand in response.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT author contributions statement

Mark Spicer: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. Jason R.C. Nurse: Methodology, Supervision, Writing - review & editing.

Data availability

The data used in this article is not available

Funding

This research was carried out as part of a PhD research project that is part funded by the PhD student's employer.

Appendices

Appendix A – Quality assessment analysis

Paper	Research Type	Sample Size	No of Authors	Citations in paper	Topics	Q1 Clarity of Aims	Q2 Appropriate Design	Q3 Risk of Bias	Q4 Dataset Realism	Q5 Sample Strength	Q6 Clarity of Findings	Total	Tier
Abroshan	Simulation	135	4	1	1	1	1	1	0.5	0	0	3.5	Medium
Albladi	Lab/Survey	316	2	1	1	1	1	0	0	0.5	0	2.5	Low
Alwanain	Simulation	3,000	1	1	3	0.5	0.5	0.5	1	0.5	0.5	3.5	Medium
Anawar	Lab/Survey	252	4	2	7	1	1	0.5	0	0.5	1	4	Medium
Bakhshi	Simulation	97	1	3	7	1	1	0.5	0.5	0	1	4	Medium
Bayl-Smith	Simulation	50,000	4	3	7	1	1	1	1	1	1	6	High
Bollens	Dataset	100	1	4	2	1	1	0.5	0.5	0	0.5	3.5	Medium
Bona	Simulation	191	2	6	6	1	1	1	0.5	0	1	4.5	Medium
Braun	Simulation	68,743	5	1	1	1	1	1	1	1	0.5	5.5	High
Champa	Lab/Survey	55	4	2	6	1	1	0.5	0	0	1	3.5	Medium
Chrysanthou	Dataset	25,205	3	1	2	1	1	0.5	0.5	1	0.5	4.5	Medium
Das	Lab/Survey	57	3	2	2	1	1	0.5	0	0	0.5	3	Low
Eftimie	Simulation	940	4	12	6	1	1	1	1	0	1	5	High
Ferreira	Lab/Survey	52	2	3	3	1	1	0	0	0	0.5	2.5	Low
Frank	Simulation	2,302	3	5	4	1	1	1	1	0.5	1	5.5	High
Gallo	Lab/Survey	500	5	2	6	1	1	0.5	0	1	1	4.5	Medium
George	Simulation	876	3	3	1	1	1	1	1	0	0.5	4.5	Medium
Gordon	Simulation	2,975,019	7	2	1	1	1	1	1	1	0.5	5.5	High
Gratian	Lab/Survey	369	5	2	6	1	1	0.5	0	1	1	4.5	Medium
Hanus	Simulation	2,866	3	2	3	1	1	1	1	0.5	0.5	5	High
Harrison	Simulation	194	3	1	2	1	1	1	1	0	0.5	4.5	Medium
Hassandoust	Simulation	N/A	3	4	3	1	1	1	1	0.5	0.5	5	High
Jalali	Simulation	430	3	7	3	1	1	1	1	0.5	1	5.5	High
Jones	Lab/Survey	68	1	2	2	0.5	0.5	0	0	0	0.5	1.5	Low
Koddebusch	Simulation	3,452	1	3	3	0.5	0.5	0.5	1	0.5	0.5	3.5	Medium
Kumar	Simulation	1,774	4	1	3	1	1	1	1	0.5	0.5	5	High
Lain (2021)	Simulation	117,864	3	1	1	1	1	1	1	1	1	6	High
Lain (2022)	Simulation	13,662	3	5	2	1	1	1	1	1	1	6	High
Lawson	Lab/Survey	102	4	3	7	1	1	0.5	0	0.5	1	4	Medium
Longtchi	Dataset	1,260	7	4	1	1	1	1	0.5	0.5	1	5	High
Luga	Lab/Survey	382	3	4	2	1	1	0.5	0	1	0.5	4	Medium
McElwee	Dataset	N/A	3	1	3	1	1	0.5	0.5	0.5	0.5	4	Medium
Moody	Lab/Survey	632	3	3	7	1	1	0.5	0.5	1	1	5	High
Musava	Simulation	4,483	3	3	3	1	1	1	1	0.5	0.5	5	High
O'Connor	Lab/Survey	196	4	0	1	1	1	0.5	0	0.5	0.5	3.5	Low
Oliveira	Lab/Survey	158	10	6	4	1	1	1	0.5	1	1	5.5	High
Oner	Simulation	195,000	3	1	4	1	1	1	1	1	1	6	High
Parsons	Lab/Survey	117	5	7	4	1	1	0.5	0	0.5	1	4	Medium
Pinto	Simulation	1,025	3	3	2	1	1	1	1	0.5	0.5	5	High
Sarno (2020)	Simulation	300	4	1	2	1	1	1	1	0	1	5	High
Sarno (2022)	Lab/Survey	69	2	1	3	1	1	0.5	0	0	1	3.5	Medium
Scott	Simulation	360	4	4	2	1	1	1	1	0	0.5	4.5	Medium
Sheng	Lab/Survey	1,001	5	5	3	1	1	1	0	1	0.5	4.5	Medium
Smith	Simulation	646	2	1	2	1	1	1	1	0	1	5	High
Steves	Simulation	9,654	3	3	3	1	1	1	1	1	1	6	High
Sutter	Simulation	288,000	4	5	2	1	1	1	1	1	1	6	High
Taib	Simulation	56,365	5	2	2	1	1	1	1	1	1	6	High
Toth	Simulation	13,000	4	11	7	1	1	1	1	1	1	6	High
Vishwanath (2011)	Simulation	345	5	8	5	1	1	1	1	1	1	6	High
Vishwanath (2018)	Simulation	161	3	1	1	1	1	1	1	1	1	6	High
Wang	Dataset	10	2	2	2	1	1	0.5	0.5	0	0.5	3.5	Medium
Williams	Simulation	125,284	3	13	8	1	1	1	1	1	1	6	High
Yang	Simulation	2,210	6	1	3	1	1	1	1	0.5	0.5	5	High
Zielinska	Dataset	887	4	1	3	1	1	0.5	0.5	0	0.5	3.5	Medium

Table 9 - Quality Assessment analysis.

Appendix B – Quality assessment scoring criteria

High	Total ≥ 5.0 — strong methodological quality
Medium	Total 3.5–4.5 — adequate methodological quality with limitations
Low	Total < 3.5 — substantive methodological limitations; findings treated with additional caution

Table 10 - Quality Tier thresholds

Criterion	Score 1.0	Score 0.5	Score 0
Q1 — Aims and research questions	Aims clearly stated; ≥ 2 authors	Single-author paper (aims may be less rigorously articulated)	Aims unclear (rare in peer-reviewed work)
Q2 — Research design	Multi-author peer-reviewed paper with appropriate methodology	Single-author paper or conference-only paper, thinner methods section	Design fundamentally inappropriate (rare in peer-reviewed work)
Q3 — Risk of bias (Field-specific)	Multi-author (≥ 3) real-world simulation with observed behavioural outcome (lowest awareness bias and lowest authorship bias)	Real-world simulation with 1–2 authors; OR multi-author lab/survey study; OR multi-author third-party dataset analysis	Single-author lab/survey study
Q4 — Dataset realism / ecological validity (Field-specific)	Real-world phishing simulation (operational deployment with observed behaviour)	Third-party phishing dataset analysis	Lab experiment/structured interview (no real-world phishing exposure)
Q5 — Sample strength (Field-specific)	Simulation $\geq 10,000$ emails; OR dataset $\geq 5,000$ items; OR lab/survey ≥ 200 participants	Simulation 1,000–9,999 emails; OR dataset 1,000–4,999 items; OR lab/survey 100–199 participants; OR sample size unavailable	Simulation $< 1,000$ emails; OR dataset $< 1,000$ items; OR lab/survey < 100 participants
Q6 — Clarity and validity of findings	Cited ≥ 6 times in this SLR OR covers ≥ 4 topics	Cited 3–5 times OR covers 2–3 topics	Cited ≤ 2 times AND covers ≤ 1 topic

Table 11 - Quality scoring criterion

References

- Abood N. Big Five traits: a critical review. ResearchGate; 2019. <https://www.researchgate.net/publication/335834113>.
- Abroshan H, Devos J, Poels G, Laermans E. Phishing happens beyond technology: the effects of human behaviours and demographics on each step of a phishing process. IEEE Access 2021;9:44928–44941. <https://doi.org/10.1109/ACCESS.2021.3066383>.
- Albladi SM, Weir GRS. Predicting individuals' vulnerability to social engineering in social networks. Cybersecurity 2020;3(1):7. <https://doi.org/10.1186/s42400-020-00047-5>.
- Alwanain MI. Effects of user-awareness on the detection of phishing emails: a case study. Int J Comput Sci Netw Secur 2019;19(2):57–62.
- Anawar S, Ibrahim MA, Zakaria NA, Abidin ZZ. Analysis of phishing susceptibility in a workplace: a Big Five personality perspectives. J Eng Sci Technol 2019;14(5):2516–2535.
- Bakhshi T. Social engineering: revisiting end-user awareness and susceptibility to classic attack vectors. Proc IEEE ICET 2017:1–6. <https://doi.org/10.1109/ICET.2017.8281653>.
- Baki S, Verma R. Sixteen years of phishing user studies: what have we learned? IEEE Transactions on Dependable and Secure Computing 2023;20(2):1200–1212. <https://doi.org/10.1109/TDSC.2022.3151103>.
- Bayl-Smith P, Taib R, Yu K, Berkovsky S. Dimensions of susceptibility to phishing. ACM Trans Priv Secur 2021;24(3):1–30. <https://doi.org/10.1145/3447815>.
- Bollens K. A practical investigation of spear phishing spam emails: comparative analysis and evaluation. Proc IEEE BigData 2024:1–8. <https://doi.org/10.1109/BigData66906.2024.10773554>.
- Bona M, Paci F. A real world study on employees' susceptibility to phishing attacks. Proc 15th Int Conf Availability, Reliability and Security (ARES '20). 2020:1–8. <https://doi.org/10.1145/3407023.3409179>.
- Braun O, Mayrhofer F, Böhm F, Payer T, Posegga J. Different seas, different phishes – large-scale analysis of phishing simulations across different industries. Proc ACM ASIACCS 2025. <https://doi.org/10.1145/3708821.3733905>.
- Bushman BJ, Wang MC. Vote-counting procedures in meta-analysis. In: Cooper H, Hedges LV, Valentine JC, editors. The Handbook of Research Synthesis and Meta-Analysis. 2nd ed. New York: Russell Sage Foundation; 2009. p. 207–220
- Carroll F, Adejobi A, Montasari R. How good are we at detecting a phishing attack? Investigating the evolving phishing attack email and why it continues to successfully deceive society. SN Comput Sci 2022;3:122. <https://doi.org/10.1007/s42979-022-01069-1>.
- Champa AI, Rabbi MF, Eishita FZ, Zibrán MF. Catch me if you see: using visual cue and explanatory feedback to enhance human phishing detection. IEEE Access 2025;13:193249–193272. <https://doi.org/10.1109/ACCESS.2025.3631670>.
- Chrysanthou A, Pantis Y, Patsakis C. The anatomy of deception: measuring technical and human factors of a large-scale phishing campaign. Comput Secur 2024;140:103780. <https://doi.org/10.1016/j.cose.2024.103780>.

- Cialdini RB. Influence: the psychology of persuasion. Rev. ed. New York: HarperCollins; 2007.
- Das A, Baki S, El Aassal A, Verma R, Dunbar A. SoK: a comprehensive reexamination of phishing research from the security perspective. *IEEE Commun Surv Tutor* 2020;22(1):671–708. <https://doi.org/10.1109/COMST.2019.2957750>.
- Das S, Nippert-Eng C, Camp LJ. Evaluating user susceptibility to phishing attacks. *Information and Computer Security* 2022;30(1):1–18. <https://doi.org/10.1108/ICS-12-2020-0204>.
- Desolda G, Ferro LS, Marrella A, Marsico MD, Vitiello S. Human factors in phishing attacks: a systematic literature review. *ACM Comput Surv* 2022;54(8):1–35. <https://doi.org/10.1145/3469886>.
- Dybå T, Dingsøyr T. Empirical studies of agile software development: a systematic review. *Inf Softw Technol* 2008;50(9–10):833–859. <https://doi.org/10.1016/j.infsof.2008.01.006>.
- Eftimie S, Timofte C-R, Moga R, Rus O. Spear-phishing susceptibility stemming from personality traits. *Proc IEEE ICCP* 2022:253–259. <https://doi.org/10.1109/ICCP56966.2022.9826749>.
- Ferreira A, Lanzini P. An analysis of social engineering principles in effective phishing. *Proc Workshop STAST* 2015:9–16. <https://doi.org/10.1109/STAST.2015.10>
- Frank M, Jaeger L, Montasari R. Contextual drivers of employees' phishing susceptibility: insights from a field study. *Decis Support Syst* 2022;157:113765. <https://doi.org/10.1016/j.dss.2022.113765>
- Gallo L, Gentile D, Ruggiero S, Botta A, Ventre G. The human factor in phishing: collecting and analysing user behaviour when reading emails. *Computers & Security*. 2024;139:103671. <https://doi.org/10.1016/j.cose.2023.103671>.
- George MS, Teunisse AK, Case TI. Gotcha! Behavioural validation of the Gullibility Scale. *Pers Individ Differ* 2020;162:110034. <https://doi.org/10.1016/j.paid.2020.110034>.
- Ghazali N, Ahmad N, Napis S, Kasim H. The evolution of research on phishing: a bibliometric analysis (2005–2025). *Int J Adv Comput Sci Appl* 2025. <https://doi.org/10.14569/IJACSA.2025.0160133>.
- Gollwitzer PM. Implementation intentions: strong effects of simple plans. *Am Psychol* 1999;54(7):493–503. <https://doi.org/10.1037/0003-066X.54.7.493>.
- Gordon WJ, Wright A, Glynn RJ, Kadakia J, Mazzone C, Leineweber E, Landman A. Assessment of employee susceptibility to phishing attacks at US health care institutions. *JAMA Netw Open* 2019;2(3):e190393. <https://doi.org/10.1001/jamanetworkopen.2019.0393>.
- GOV.UK. Cyber security breaches survey 2025 [Internet]. 2025 [cited 2025 Sep 5]. Available from: <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2025/cyber-security-breaches-survey-2025>.
- Gratian M, Bandi S, Cukier M, Dykstra J, Ginther A. Correlating human traits and cyber security behavior intentions. *Comput Secur* 2018;73:345–358. <https://doi.org/10.1016/j.cose.2017.11.015>.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). McGraw-Hill.

Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 8th ed. Andover: Cengage Learning; 2018.

Hanus B, Wu YA, Parrish J. Phish me, phish me not. *J Comput Inf Syst* 2022;62(3):516–526. <https://doi.org/10.1080/08874417.2020.1858730>.

Harrison B, Vishwanath A, Rao R. A user-centered investigation of personal email security. *Proc HICSS* 2016:5929–5937. <https://doi.org/10.1109/HICSS.2016.733>.

Hasegawa A, Hiraki T, Yamada S, Asai Y, Kasahara S, Yamashita T, Mori T. How WEIRD is usable privacy and security research? In: 33rd USENIX Security Symposium (USENIX Security 24); 2024; Philadelphia, PA, USA. p. 3621–3638.

Hassandoust F, Singh H, Williams J. The role of contextualization in individuals' vulnerability to phishing attempts. *Australas J Inf Syst* 2020;24. <https://doi.org/10.3127/ajis.v24i0.2693>.

Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? *Behav Brain Sci* 2010;33(2-3):61–83. <https://doi.org/10.1017/S0140525X0999152X>.

Hofstede G, Hofstede GJ, Minkov M. *Cultures and organizations: software of the mind*. 3rd ed. New York: McGraw-Hill; 2010.

Human Factors and Ergonomics Society. What is human factors and ergonomics? HFES; 2024. <https://www.hfes.org/About/What-Is-Human-Factors-and-Ergonomics>.

Jabir R, Le J, Nguyen C. Phishing attacks in the age of generative artificial intelligence: a systematic review of human factors. *AI* 2025;6(8):174. <https://doi.org/10.3390/ai6080174>.

Jalali MS, Siegel M, Madnick S. Decision-making and biases in cybersecurity capability development: evidence from a simulation game experiment. *J Strateg Inf Syst* 2019;28(1):66–82. <https://doi.org/10.1016/j.jsis.2018.09.003>.

Jones HS. What makes people click: assessing individual differences in susceptibility to email fraud [dissertation]. [Bournemouth]: Bournemouth University; 2016.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kitchenham B, Brereton P, Li Z, Budgen D, Burn A. Repeatability of systematic literature reviews. *Proc EASE* 2011. <https://doi.org/10.1049/ic.2011.0006>.

Kitchenham B, Charters S. Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01. Keele University and Durham University; 2007.

Koddebusch M. Exposing the phish: the effect of persuasion techniques in phishing emails. *Proc EICC '22*. 2022:1–6. <https://doi.org/10.1145/3543434.3543476>.

Kumar A, Gupta S, Rai AK, Kumar S. Analysing user vulnerability to phishing attacks using machine learning techniques: a case study on university data. *Proc ICICAT* 2023. <https://doi.org/10.1109/ICICAT57735.2023.10487734>.

Lain D, Kostianen K, Čapkun S. Phishing in organizations: findings from a large-scale and long-term study. *Proc ACM CCS '21*. 2021:1566–1583. <https://doi.org/10.1145/3460120.3484590>.

- Lain D, Kostianen K, Čapkun S. Content, nudges and incentives: a study on the effectiveness and perception of embedded phishing training. *Proc IEEE S&P* 2022;1–18. <https://doi.org/10.1109/SP46214.2022.9833571>.
- Lawson P, et al. Email phishing and signal detection: how persuasion principles and personality influence response patterns and accuracy. *Appl Ergon* 2020;86:103084. <https://doi.org/10.1016/j.apergo.2020.103084>.
- Longtchi TT, Montañez Rodriguez R, Gwartney K, Ear E, Azari DP, Kelley CP, Xu S. Quantifying psychological sophistication of malicious emails. *IEEE Access* 2024;12:187512–187535. <https://doi.org/10.1109/ACCESS.2024.3510534>.
- Luga C, Seigneur J-M, Launois H. Baiting the hook: factors impacting susceptibility to phishing attacks. In: *Proceedings of HAISA 2016*; 2016; Frankfurt, Germany. p. 1–10.
- Marin I, Burda P, Allodi L, Zannone N. The influence of human factors on the intention to report phishing emails. *Proc ACM CHI '23*. 2023;1–18. <https://doi.org/10.1145/3544548.3580985>.
- McElwee S, Murphy G, Shelton P. Influencing outcomes and behaviors in simulated phishing exercises. *Proc IEEE SoutheastCon* 2018;1–6. <https://doi.org/10.1109/SECON.2018.8479109>.
- Memon, M. A., Ting, H., Cheah, J. H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), i–xx. [https://doi.org/10.47263/JASEM.4\(2\)01](https://doi.org/10.47263/JASEM.4(2)01).
- Moody GD, Galletta DF, Dunn BK. Which phish get caught? An exploratory study of individuals' susceptibility to phishing. *Eur J Inf Syst* 2017;26(6):564–584. <https://doi.org/10.1057/s41303-017-0058-x>.
- Musava P, Labuschagne N, Veerasamy N. A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Comput Human Behav* 2019;94:357–369. <https://doi.org/10.1016/j.chb.2018.12.036>.
- Naqvi B, Perova K, Farooq A, Makhdoom I, Oyedele S, Porras J. Mitigation strategies against the phishing attacks: a systematic literature review. *Comput Secur* 2023;132:103387. <https://doi.org/10.1016/j.cose.2023.103387>.
- NCSC. Phishing attacks: defending your organisation. 2024. <https://www.ncsc.gov.uk/guidance/phishing>.
- O'Connor AM, Judges RA, Lee K, Evans AD. Can adults discriminate between fraudulent and legitimate e-mails? Examining the role of age and prior fraud experience. *J Elder Abuse Negl* 2021;33(3):181–205. <https://doi.org/10.1080/08946566.2021.1934767>.
- Oliveira D, Rocha H, Yang H, Ellis D, Dommaraju S, Muradoglu M, Weir D, Soliman A, Lin T, Ebner N. Dissecting spear phishing emails for older vs young adults: on the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*; 2017; Denver, CO, USA. New York: ACM; p. 6412–6424. <https://doi.org/10.1145/3025453.3025831>.
- Oner U, Cetin O, Savas E. Human factors in phishing: understanding susceptibility and resilience. *Comput Stand Interfaces* 2025;94:104014. <https://doi.org/10.1016/j.csi.2025.104014>.

Owoh BR, Haig E. A systematic review of voluntary cybersecurity standards and frameworks. *Int J Inf Secur* 2025;24:65. <https://doi.org/10.1007/s10207-025-01121-0>.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.

Parsons K, McCormac A, Pattinson M, Butavicius M, Jerram C. Phishing for the truth: a scenario-based experiment of users' behavioural response to emails. *Proc IFIP SEC* 2013:371–378. https://doi.org/10.1007/978-3-642-39218-4_27.

Patterson C, Nurse JRC, Franqueira VNL. Learning from cyber security incidents: a systematic review and future research agenda. *Comput Secur* 2023;132:103309. <https://doi.org/10.1016/j.cose.2023.103309>.

Petric G, Just JN. Information security culture and phishing-reporting model: structural equivalence across Germany, UK, and USA. *J Cybersecur* 2025;11(1):tyaf011. <https://doi.org/10.1093/cybsec/tyaf011>.

Pinto L, Filiol E, Cuppens F. Assessing the relevance of cybersecurity training and policies to prevent and mitigate the impact of phishing attacks. *J Inf Syst Secur* 2022;18(2). <https://jisis.org/wp-content/uploads/2022/12/14.002.pdf>.

Ribeiro L, Guedes IS, Cardoso CS. Which factors predict susceptibility to phishing? An empirical study. *Comput Secur* 2024;136:103558. <https://doi.org/10.1016/j.cose.2023.103558>.

Rothstein HR, Sutton AJ, Borenstein M, editors. *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester: Wiley; 2005.

Salloum SA, Gaber T, Vadera S, Shaalan K. A systematic literature review on phishing email detection using natural language processing techniques. *IEEE Access* 2022;10:65703–65727. <https://doi.org/10.1109/ACCESS.2022.3183083>.

Sarno DM, Lewis JE, Bohil CJ, Neider MB. Which phish is on the hook? Phishing vulnerability for older versus younger adults. *Hum Factors* 2020;62(5):704–717. <https://doi.org/10.1177/0018720819855570>.

Sarno DM, Neider MB. So many phish, so little time: exploring email task factors and phishing susceptibility. *Hum Factors* 2022;64(8):1379–1403.

Scott J, Levy Y, Li W, Kumar A. Comparing phishing training and campaign methods for mitigating malicious emails in organizations. *Proc ICISSP* 2024:643–651. <https://doi.org/10.5220/0012281600003648>.

Sheng S, Holbrook M, Kumaraguru P, Cranor LF, Downs J. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proc ACM CHI '10*. 2010:373–382. <https://doi.org/10.1145/1753326.1753383>.

Smith SE, Dupuis MJ. Phishing experiments in the wild: lessons for ubiquitous and context-aware security. *Proc IEEE UEMCON* 2025:647–654. <https://doi.org/10.1109/UEMCON67449.2025.11267764>.

- Steinberg J. Cybersecurity for dummies. 2nd ed. Hoboken: Wiley; 2022.
- Steves M, Greene K, Theofanos M. Categorizing human phishing difficulty: a Phish Scale. *J Cybersecur* 2020;6(1):tyaa009. <https://doi.org/10.1093/cybsec/tyaa009>.
- Sutter T, Distler V, Lallemand C, Koenig V. Avoiding the hook: influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. *Proc IEEE EuroS&PW* 2022:35–44. <https://doi.org/10.1109/EuroSPW55150.2022.00012>.
- Taib R, Yu K, Berkovsky S, Wiggins M, Bayl-Smith P. Social engineering and organisational dependencies in phishing attacks. *Proc INTERACT* 2019:564–573. https://doi.org/10.1007/978-3-030-29381-9_34.
- Tedja B, Manafe L, Nursalam N, Kristianto H. Systematic literature review using PRISMA: exploring the influence of service quality and perceived value on satisfaction and intention to continue relationship. *Futur Bus J* 2024;10:27. <https://doi.org/10.1186/s43093-024-00326-4>.
- Tornblad MK, Jones KS, Namin AS, Choi J. Characteristics that predict phishing susceptibility: a review. *Proc Hum Factors Ergon Soc Annu Meet* 2021;65(1):938–942. <https://doi.org/10.1177/1071181321651330>.
- Toth R, Dubniczky RA, Limonova O, Tihanyi N. Sustaining cyber awareness: the long-term impact of continuous phishing training and emotional triggers. *Proc IEEE BigData* 2025:7854–7862. <https://doi.org/10.1109/BigData66926.2025.11402180>.
- Uchendu B, Nurse JRC, Bada M, Furnell S. Developing a cyber security culture: current practices and future needs. *Comput Secur* 2021;109:102387. <https://doi.org/10.1016/j.cose.2021.102387>.
- Vishwanath A, Herath T, Chen R, Wang J, Rao HR. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decis Support Syst* 2011;51(3):576–586. <https://doi.org/10.1016/j.dss.2011.03.002>.
- Vishwanath A, Harrison B, Ng YJ. Suspicion, cognition, and automaticity model of phishing susceptibility. *Commun Res* 2018;45(8):1146–1166. <https://doi.org/10.1177/0093650215627483>.
- Wang P, Lutchkus P. Psychological tactics of phishing emails. *Issues Inf Syst* 2023;24(2):71–83.
- Washo AH. An interdisciplinary view of social engineering: a call to action for research. *Comput Hum Behav Rep* 2021;4:100126. <https://doi.org/10.1016/j.chbr.2021.100126>.
- Williams EJ, Hinds J, Joinson AN. Exploring susceptibility to phishing in the workplace. *Int J Hum Comput Stud* 2018;120:1–13. <https://doi.org/10.1016/j.ijhcs.2018.06.004>.
- Wood W, Neal DT. A new look at habits and the habit-goal interface. *Psychol Rev* 2007;114(4):843–863. <https://doi.org/10.1037/0033-295X.114.4.843>.
- Yang R, Zheng K, Wu B, Li D, Wang Z, Wang X. Predicting user susceptibility to phishing based on multidimensional features. *Comput Intell Neurosci* 2022;2022:7058972. <https://doi.org/10.1155/2022/7058972>.

Zielinska OA, Welk AK, Mayhorn CB, Murphy-Hill E. A temporal analysis of persuasion principles in phishing emails. *Proc Hum Factors Ergon Soc Annu Meet* 2016;60(1):765–769. <https://doi.org/10.1177/1541931213601175>.

Zhuo S, Biddle R, Koh YS, Lottridge D, Russello G. SoK: human-centered phishing susceptibility. *ACM Trans Priv Secur* 2023;26(3):24. <https://doi.org/10.1145/3575797>.