HUMAN FACTORS IN CYBERSECURITY: FROM HUMAN ERROR TO SOCIO-TECHNICAL RESILIENCE

CZU: 005.331:004.056
DOI: https://doi.org/10.53486/tids2026.33

EVELINA STEBNITSKAIA
Financial University under the Government of Russian Federation
e-mail: thestebev@gmail.com
ORCID ID: 0009-0006-5279-8255

DARIA KOZLOVA
Financial University under the Government of Russian Federation
e-mail: dasha280206k@gmail.com
ORCID ID: 0009-0009-6981-0259

DINARA R. ORLOVA
DSC, Professor, Department of Economics
Financial University under the Government of the Russian Federation
e-mail: DROrlova@fa.ru
ORCID ID: 0000-0002-2901-070X

Abstract. This paper examines human factors in cybersecurity as structural drivers of cyber risk rather than as residual deviations from otherwise secure technical systems. Although cybersecurity discourse still often describes users as the weakest link, contemporary research shows that breaches are more accurately produced through the interaction of cognitive limits, behavioural tendencies, interface design, organisational climate and adversarial manipulation of context. The paper aims to analyse the main cognitive, behavioural and organisational human factors that intensify cyber vulnerability and to formulate mitigation strategies that reduce risk without undermining legitimate work practices. Methodologically, the study adopts a critical integrative literature review that synthesises peer-reviewed research published between 2018 and 2025 together with foundational work in human error theory and behavioural decision-making. The analysis is organised around seven domains: human error, cognitive bias, phishing susceptibility, authentication behaviour, organisational security climate, training effectiveness and human-centred design. The review demonstrates that users do not fail simply because they are careless. Slips, lapses and mistakes are amplified by fatigue, divided attention, ambiguous interfaces, time pressure and contradictory productivity incentives. Phishing succeeds not only because messages are deceptive, but because recipients interpret cues through work context and fast, heuristic processing. Authentication failures likewise arise from the security-usability trade-off that encourages password reuse, approval fatigue and informal workarounds. At the organisational level, leadership visibility, procedural justice, policy congruence and peer norms improve voluntary compliance more reliably than fear-based messaging alone. The paper argues that resilient cybersecurity requires a shift from a blame-oriented paradigm to a socio-technical model that treats users as adaptive actors who must be supported by secure systems, credible governance and continuous learning. It concludes that the strongest approach combines human-risk assessment, phishing-resistant authentication, repeated context-aware training, forgiving interfaces, behavioural metrics and no-blame reporting.

Keywords: human factors in cybersecurity, socio-technical security, phishing susceptibility, security climate, authentication behaviour, human-centred design

JEL Classification: D91, M15, O33

INTRODUCTION

Cybersecurity breaches are often narrated as technical failures, yet the decisive moment in many incidents is behavioural rather than purely computational. A malicious email is opened, a credential is reused, an alert is dismissed, a policy is bypassed for the sake of continuity, or a manager delays remediation because the threat appears abstract. These episodes are not peripheral to cybersecurity; they are central to it. The latest Verizon DBIR executive summary still reports that the human element is implicated in roughly 60 per cent of analysed breaches, which confirms that cyber risk is generated at the intersection of human conduct, organisational incentives and technical architecture (Verizon, 2025).

The core problem addressed in this paper is the persistent gap between formal security policy and everyday user behaviour. Organisations publish password rules, mandate secure data handling, deploy multi-factor authentication and conduct awareness campaigns, yet employees still click on phishing links, share credentials, disable protections, or invent workarounds that restore convenience at the expense of resilience. This gap cannot be understood adequately through a purely technical lens because policy non-compliance is frequently rooted in overload, ambiguity, competing goals, weak security climate, or poor usability rather than deliberate negligence alone.

Recent literature illuminates different parts of this puzzle. Ncubukezit (2022) shows how recurrent human errors undermine small-business security. Barre et al. (2025) examine the role of cognitive biases in cyber-risk communication between boards and CISOs. Greene *et al.* (2018) demonstrate that phishing susceptibility is strongly shaped by work context rather than message content alone. Gwebu and Wang (2024) analyse the defining features of a robust information security climate. Aldawood and Skinner (2019) explain why many awareness programmes underperform, while Zimmermann *et al.* (2024) challenge the deeply rooted image of users as enemies rather than partners. Read together, these studies suggest a common theoretical proposition: cyber vulnerability is produced by socio-technical interaction, not by user defect in isolation.

Accordingly, the aim of this paper is to analyse how cognitive, behavioural and organisational human factors generate cybersecurity vulnerability and to propose academically grounded mitigations. The objectives are fourfold: to identify recurrent human vulnerabilities in contemporary cyber incidents; to evaluate the explanatory value of human error theory, bounded rationality and organisational climate perspectives; to assess the strengths and limits of training and authentication interventions; and to provide recommendations for human-centred cyber resilience. The originality of the paper lies in integrating these strands into a single analytical framework that treats human behaviour not as an afterthought to technical security, but as one of its governing conditions.

MAIN CONTENT

1. Materials and Methods

This study uses a critical integrative literature review as its principal method. Rather than merely listing prior findings, the review compares explanatory mechanisms across bodies of research and develops a conceptual synthesis. The core materials consist of peer-reviewed journal articles, conference papers, institutional publications and one carefully selected academic thesis published mainly between 2018 and 2025, supplemented by foundational works on human error and judgement under uncertainty.

The literature was selected according to thematic relevance to seven analytical domains: user error, cognitive bias, phishing and social engineering, authentication adoption, organisational security

climate, training effectiveness and human-centred security design. Analytically, the sources were organised into three interacting levels. The first level is cognitive and concerns attention, memory, heuristics and bias. The second is behavioural and concerns situated action, routine, impulsivity and compliance under workload. The third is organisational and concerns leadership, norms, incentives, procedural justice and the design of work systems. This multilevel perspective is necessary because the same failure event may be interpreted differently depending on the unit of analysis.

The method has two limitations that should be acknowledged. First, the study is a conceptual synthesis rather than a primary empirical investigation, so its contribution is interpretive and prescriptive rather than statistically causal. Secondly, the reviewed studies employ heterogeneous designs, including literature reviews, field experiments, organisational surveys, conference papers and conceptual articles, which means that direct comparison requires analytical judgement. Nevertheless, this diversity is also a strength, because human factors in cybersecurity are irreducibly interdisciplinary and cannot be captured through a single methodological lens.

2. Human Error as a Structural Cybersecurity Variable

Human error remains one of the most widely cited explanations for cyber incidents, but the concept is often used too loosely. In classical human factors research, error is not a homogeneous category. Reason (1990) distinguishes among slips, lapses and mistakes. Slips are failures of execution despite a correct intention; lapses are failures of memory or recall; and mistakes arise when the intention itself is flawed because the actor applies the wrong rule, mental model or judgement. This taxonomy is especially valuable for cybersecurity because it differentiates interface-induced action failures, memory burden and deeper failures of understanding.

Ncubukezit (2022) argues that recurrent human errors constitute a major cybersecurity concern for small businesses and identifies ignorance, poor decision-making, technical errors and skills- or policy-related failures as recurring sources of vulnerability. The broader implication is that error is rarely random. Users send information to the wrong recipient, misconfigure permissions, ignore software updates, or approve suspicious requests within environments defined by interruptions, multitasking, weak feedback and contradictory deadlines. Contemporary breach reporting supports this structural interpretation: even under a relatively strict coding framework, Verizon (2025) still attributes about 60 per cent of analysed breaches to the human element.

For this reason, the phrase weakest link is analytically insufficient. It individualises what is often a system output. Fatigue, alert saturation, divided attention and constant context switching make slips and lapses more probable, while poorly translated policies and weak workflow support make mistakes more likely. A rigorous approach to human error in cybersecurity therefore asks not only who erred, but what properties of the task, interface, workflow and governance made the error foreseeable.

3. Cognitive Biases, Heuristics and Risk Perception

Human behaviour in cybersecurity is shaped not only by error in the narrow sense, but also by ordinary patterns of judgement under uncertainty. Kahneman (2011) describes the fast, intuitive operations of System 1 and the slower, reflective operations of System 2. In routine organisational life, System 1 often dominates because people must act quickly, triage information and preserve cognitive resources. The consequence is not irrationality in an absolute sense, but bounded rationality: people make workable decisions under constraint, often at the cost of systematic bias.

In cybersecurity, optimism bias encourages the belief that one is less likely than others to become a victim. Confirmation bias leads users and managers alike to privilege cues that fit existing

expectations, such as assuming that a familiar-looking invoice or routine service notice is legitimate. Status quo bias encourages employees and executives to preserve insecure but familiar practices because change introduces friction, learning costs and uncertainty. Barre *et al.* (2025) show that such biases are especially consequential in cybersecurity governance, where boards and CISOs must interpret uncertain risk signals and communicate them across professional boundaries. Distorted risk perception at this level can lead to underinvestment, symbolic compliance and dangerous overconfidence.

The significance of cognitive bias lies in its cross-level character. The same mechanisms that help an employee process inbox traffic quickly can also cause a board to minimise abstract cyber exposure if no dramatic incident has recently occurred. Bias is therefore not a problem of end users only. It permeates the whole governance chain, from frontline message handling to strategic decision-making. Effective mitigation requires interventions that slow high-stakes decisions, improve the salience of risk-relevant cues and reduce the need for users to infer security significance from ambiguous signals.

4. Phishing as a Context-Sensitive Behavioural Attack

Phishing is frequently discussed as an issue of message deception, yet the evidence shows that it is more accurately understood as a context-sensitive behavioural attack. Greene *et al.* (2018), drawing on 4.5 years of embedded phishing exercises in a U.S. government research environment, found that user context was a significant explanatory variable in phishing susceptibility. Their key contribution is not simply that some users click more than others, but that clickers and non-clickers attend to different cues and interpret the same cues differently depending on how closely a message aligns with ongoing work tasks.

This finding has far-reaching implications. Urgency does not work in isolation; it works because it resonates with organisational expectations of responsiveness. Likewise, polished visual design, familiar senders, invoice language, delivery notifications and workflow plausibility all operate through situated interpretation. Phishing exploits normal organisational routines rather than isolated cognitive defects. Susceptibility therefore increases under time pressure, cognitive overload and cultures in which speed is rewarded more visibly than verification.

Simulated phishing campaigns remain useful, but only when they are treated as diagnostic instruments rather than tools of humiliation. Their principal value lies in identifying high-risk contexts, communication patterns and role-specific vulnerabilities. Used well, they reveal whether certain teams are structurally exposed because of job design, customer-facing responsibilities or dense transactional communication. Phishing defence should therefore move beyond the generic lesson to be careful and toward the redesign of cues, workflows and escalation paths so that suspicious requests become easier to pause, verify and report.

5. Authentication, Memory Burden and MFA Fatigue

Authentication is a classic site of tension between security and usability. Password systems rely heavily on human memory, yet strong password guidance typically demands length, uniqueness, complexity and regular management across many accounts. The predictable result is password reuse, insecure storage practices, low-quality coping strategies and eventual fatigue. In practice, users do not reject security because they oppose it in principle; they reject authentication regimes that impose excessive cognitive or temporal cost for uncertain benefit.

Karuppiah (2025) synthesises behavioural, usability and organisational factors influencing MFA adoption and shows that acceptance depends not only on security awareness, but also on

workflow disruption, recovery design, rigidity of enforcement and perceived fairness of rollout. Marky *et al.* (2022) likewise demonstrate that many users view two-factor authentication as necessary but annoying: they accept the protective rationale while simultaneously experiencing friction, interruption and resentment. This helps explain why approval fatigue, prompt habituation and informal bypass behaviour persist even after organisations deploy technically stronger controls.

The implication is not that authentication should be relaxed, but that it should be redesigned around human capability. Password managers reduce memory burden and password reuse. Biometrics and device-based authentication can improve convenience when implemented with careful privacy and recovery safeguards. Most importantly, phishing-resistant approaches such as syncable authenticators or passkeys can reduce dependence on manual code entry and shared secrets. NIST guidance explicitly recognises that properly configured syncable authenticators can support AAL2 while advancing broader adoption of phishing-resistant authentication (Galluzzo *et al.*, 2024). The strategic lesson is clear: the most resilient authentication systems are those that strengthen security by removing unnecessary human strain.

6. Organisational Security Climate and Voluntary Compliance

Cybersecurity behaviour is not produced by individual psychology alone; it is also shaped by the normative environment of the organisation. Gwebu and Wang (2024) define a robust information security climate as one in which security expectations are understandable, supported, enacted consistently and integrated into daily work. In such settings, employees perceive that secure behaviour is genuinely expected, practically enabled and fairly evaluated.

This perspective is important because procedural justice matters. Users are more likely to comply voluntarily when they believe policies are reasonable, leadership models the same standards and reporting channels are credible. Conversely, when management celebrates productivity while informally rewarding policy circumvention, staff learn that security is rhetorical rather than real. The organisational climate then teaches employees to treat policy as a symbolic layer to be worked around whenever deadlines, customer pressure or convenience demand it.

A strong security climate therefore depends on leadership visibility, policy congruence, peer norms and support resources. It is not enough to issue directives. Secure behaviour becomes stable only when the organisation reduces conflict between operational success and policy compliance. This is why mature security culture is less about slogans than about governance design: incentives, managerial conduct, reporting practices and everyday workflow must all signal that security is a shared institutional norm rather than a private burden placed on users.

7. Training Effectiveness Beyond Awareness Rituals

Security training is one of the most widespread responses to human cyber risk, yet its effectiveness is often overstated. Aldawood and Skinner (2019) show that many social engineering awareness programmes suffer from familiar weaknesses: limited budgets, weak tailoring, insufficient reinforcement and poor alignment with the changing tactics of attackers. One-time annual training may satisfy audit expectations, but it rarely produces durable behavioural change.

The problem is partly pedagogical and partly contextual. People forget abstract warnings when those warnings are detached from real tasks, real incentives and real moments of vulnerability. Repeated, interactive and role-specific interventions are more effective because they embed learning in practice rather than in symbolic compliance rituals. Training also works better when it is

immediately connected to suspicious events, simulated exercises or just-in-time warnings that appear inside the workflow where a risky decision is actually being made.

Training outcomes are further moderated by differences in digital literacy, professional role and prior experience. A homogeneous programme delivered to heterogeneous users will predictably underperform. This does not mean awareness initiatives are futile. It means they should be reframed as continuous learning systems that support recognition, reporting and recovery. The relevant metric is not whether employees attended a session, but whether they become more capable of noticing, pausing, escalating and learning from suspicious events over time.

8. Human-Centred Cybersecurity Design

Human-centred cybersecurity begins from the premise that people are not merely hazards to be constrained, but also partners whose capabilities can be cultivated. Zimmermann *et al.* (2024) argue that the traditional view of users as enemies has encouraged constraining approaches that limit action without adequately addressing the social and organisational conditions of secure behaviour. Their alternative is an enabling approach that treats humans as resources for resilience rather than residual sources of failure.

In design terms, this means reducing unnecessary friction, preserving user autonomy where possible and making security cues meaningful at the point of action. Warnings should communicate why a risk matters, not simply block action with generic alarm language. Interfaces should preserve evidence of origin, make verification easier than improvisation and support recovery when users make mistakes. Secure design is forgiving design: it assumes fallibility and builds in pathways for correction, escalation and learning.

This approach does not reject technical controls. It repositions them within a broader socio-technical strategy. Minimal-friction defaults, meaningful feedback, just-in-time prompts, safer recovery flows and behavioural nudges can all improve outcomes without treating the user as a problem to be neutralised. The deeper contribution of human-centred design is normative: it shifts cybersecurity from a logic of suspicion toward a logic of support.

9. Recommendations for Practice and Research

The synthesis developed in this paper supports five strategic recommendations. First, organisations should conduct human-risk assessments that examine roles, workflows and interfaces rather than relying solely on compliance checklists. The central question should be where security depends on hurried interpretation, memory load or ambiguous signals, because these are precisely the points at which slips, lapses and bias become likely. Secondly, organisations should redesign high-risk interfaces according to forgiving and human-centred principles. Security controls should communicate clearly, preserve provenance information and reduce unnecessary dependence on recall and manual verification.

Thirdly, authentication should migrate toward low-memory and phishing-resistant solutions, including password managers, well-designed MFA, resilient recovery flows and passkey-based approaches where appropriate. Fourthly, training should become continuous, role-sensitive and behaviourally measured. The goal is not symbolic awareness, but demonstrable improvement in detection, reporting and recovery. Fifthly, organisations should institutionalise no-blame reporting for near misses, suspicious events and honest mistakes. Harper (2023) is especially useful here in linking human-based cyber behaviour to practical mitigation: learning is suppressed when users

expect ridicule or punishment, yet early reporting is exactly what allows small failures to be contained before they become major incidents.

For researchers, the field now needs more longitudinal and interdisciplinary work. Future studies should compare how cognitive bias, interface design and security climate interact over time, test whether behavioural metrics can be collected ethically and validly, and examine how artificial intelligence changes both attack persuasion and defensive support. The most promising agenda is one that moves beyond the binary opposition between user failure and technical control and instead studies cybersecurity as a dynamic ecology of perception, workload, authority, design and adaptation.

CONCLUSIONS

This paper has argued that human factors in cybersecurity should be understood as a central domain of analysis rather than as a residual explanation for failures that technology did not prevent. Human error, cognitive bias, phishing susceptibility, authentication fatigue, organisational climate and training effectiveness are not separate problems; they are interconnected expressions of how people act in complex digital environments under constraint. The common lesson across the literature is that insecure behaviour is frequently predictable and often induced by the structure of work itself.

The significance and originality of the paper lie in bringing together human error theory, behavioural decision theory and organisational climate research into a unified account of cyber resilience. This synthesis helps explain why technical controls underperform when they are layered onto cognitively overloaded users and weak organisational systems. It also supports a more mature normative position: resilient cybersecurity does not arise from demanding flawless behaviour, but from designing institutions and interfaces that make secure action credible, legible and sustainable.

The paper has several limitations. It is based on conceptual synthesis rather than primary data collection, and some reviewed studies are context-specific, especially those focused on particular organisations or sectors. Future research should therefore test the proposed framework through cross-sector longitudinal studies, experimental work on warning design and organisational field studies on security climate, reporting culture and authentication behaviour. Even so, the current evidence is sufficient to reject simplistic blame narratives. In cybersecurity, the human factor is not merely the weakest link; properly understood and properly supported, it is also one of the strongest levers for resilience.

REFERENCES

Aldawood, H. and Skinner, G. (2019) ‘Reviewing cyber security social engineering training and awareness programs-pitfalls and ongoing issues’, *Future Internet*, 11(3), p. 73.

Barre, G., Huygh, T., Nguyen, D.K. and Nuijten, A. (2025) ‘Towards understanding cognitive biases in cybersecurity governance’, in Pucihar, A., Kljajić Borštnar, M., Blatnik, S., Marolt, M., Bons, R.W.H., Smit, K. and Glowatz, M. (eds.) *38th Bled eConference: Empowering Transformation: Shaping Digital Futures for All: Conference Proceedings*. Maribor: University of Maribor Press, pp. 737–744.

Galluzzo, R., Regenscheid, A., Temoshok, D. and LaSalle, C. (2024) *Incorporating Syncable Authenticators Into NIST SP 800-63B: Digital Identity Guidelines - Authentication and Lifecycle Management*. Gaithersburg, MD: National Institute of Standards and Technology.

Greene, K.K., Steves, M.P., Theofanos, M. and Kostick, J.A. (2018) ‘User context: an explanatory variable in phishing susceptibility’, in *Proceedings of the Network and Distributed System Security Symposium, Workshop on Usable Security*. San Diego, CA, 18 February 2018. Reston.

Gwebu, K.L. and Wang, J. (2024) 'The defining features of a robust information security climate', *Computers & Security*, 142, 103891.

Harper, J.W. (2023) 'Cybersecurity: a review of human-based behavior and best practices to mitigate risk', *Issues in Information Systems*, 24(4), pp. 247–254.

Kahneman, D. (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Karuppiah, S.P. (2025) *Understanding the behaviour of business users in multi-factor authentication adoption*. Master's thesis. Lappeenranta–Lahti University of Technology LUT.

Marky, K., Ragozin, K., Chernyshov, G., Matviienko, A., Schmitz, M., Mühlhäuser, M., Eghtebas, C. and Kunze, K. (2022) 'Nah, it's just annoying!' A deep dive into user perceptions of two-factor authentication, *ACM Transactions on Computer-Human Interaction*, 29(5), pp. 1–32.

Ncubukezit, T. (2022) 'Human errors: a cybersecurity concern and the weakest link to small businesses', *International Conference on Cyber Warfare and Security*, 17(1), pp. 395–403.

Reason, J. (1990) *Human Error*. Cambridge: Cambridge University Press.

Verizon (2025) *2025 Data Breach Investigations Report: Executive Summary*.

Zimmermann, V., Schöni, L., Schaltegger, T., Ambuehl, B., Knieps, M. and Ebert, N. (2024) 'Human-centered cybersecurity revisited: from enemies to partners', *Communications of the ACM*, 67(11), pp. 72–81.