WHITE PAPER

NOVEMBER 2026

Beyond Awareness

Why Cybersecurity Training Must Become Behaviour-Led

A Framework for Human-Centred Cybersecurity
Using the Cyber Rebels Five-Domain Model

Andrew Longhurst

Cyber Rebels | Revised edition, November 2026

DIGITAL OBJECT IDENTIFIER

https://doi.org/10.5281/zenodo.19039889

Executive Summary

Cybersecurity is a routine concern for organisations of different sizes. Ransomware, data breaches, online fraud and social engineering can involve emails, calls, shared documents and the digital systems people use during ordinary work. The UK Government's Cyber Security Breaches Survey provides evidence of reported exposure, although it does not measure how much relevant knowledge individual employees possess [1].

Many organisations use cybersecurity awareness training to explain risks, introduce security policies and help people recognise suspicious requests. Programmes vary considerably: some emphasise information and quizzes, while others include rehearsal, simulations and evaluation of behaviour [2, 3].

An information-only approach can rest on an appealing expectation: if people know about a cyber threat, they will avoid it. Knowledge is useful, but that expectation leaves out the circumstances in which an action must be taken.

Studies of phishing training in organisations, including interventions embedded in ordinary work, have found limited or inconsistent improvements in the actions they measured. This raises a practical question for any programme: can people apply what they know when the decision arises [4, 5]?

> This suggests that the challenge organisations face is not simply a lack of knowledge.

As explored in *Where Awareness Fails: Why Cybersecurity Training Isn't Stopping Breaches* [6], consequential cyber decisions may arise amid normal business activity: someone is handling a workload, responding to a familiar colleague or trying to keep a time-sensitive process moving. The practical issue is not only whether the person recognises a technical warning sign, but whether the task, expectations and available controls allow an appropriate response [7, 8].

Under these circumstances, knowing about a risk does not guarantee a secure decision. Knowledge, judgement, motivation, available options and the organisation's working conditions all contribute to what someone can do when a request is plausible but uncertain [9].

This paper builds on that foundation by examining what must change.

Alongside awareness, organisations need ways to develop and assess the capabilities people use while working: recognising contextual risk, completing appropriate verification, using secure processes and escalating uncertainty. Those capabilities also depend on usable tools, clear responsibilities and permission to take the safer route [3, 10].

To organise those capabilities, this paper introduces the Cyber Rebels Five-Domain Model. It is a practitioner-developed, research-informed framework for learning and assessment built around decisions made during ordinary work. Its five domains are:

- Contextual Risk Recognition
- Verification and Control Discipline
- Secure Operational Behaviour
- Incident Judgement and Escalation
- Professional Cyber Judgement

Together, these domains connect recognition, verification, secure action, escalation and professional judgement. They also provide a way to distinguish knowing the appropriate response from applying it during work, an important distinction in training-transfer research [11].

The opportunity is to develop these capabilities alongside usable controls, clear leadership expectations and ways to assess what people actually do when a decision matters.

1. The Promise of Cybersecurity Awareness Training

Cybersecurity awareness and learning programmes are established parts of organisational security practice. Their content may cover phishing, malicious attachments, credential handling, data protection and social engineering. Published evaluations show considerable variation in methods, audiences and outcomes, so “awareness training” should not be treated as a single uniform intervention [2].

Institutional approaches cover different responsibilities. NIST’s learning guidance explicitly addresses role-specific needs, behaviour and evaluation; its Cybersecurity Framework places awareness and training alongside other security outcomes. The UK’s Cyber Essentials scheme, by contrast, defines five technical control themes rather than a general requirement for awareness courses. These distinctions matter when interpreting what an organisation has demonstrated through training or certification [3, 12, 13].

As a result, awareness training became embedded within organisational security practices across both public and private sectors.

At its core, cybersecurity awareness training was built on a simple and compelling promise: that improving employees’ understanding of cyber threats would reduce the likelihood of successful attacks.

The assumption was that if employees could recognise common attack techniques – such as phishing emails, fraudulent requests or suspicious links – they would be less likely to fall victim to them. By increasing knowledge of cyber risks and reinforcing organisational security policies, awareness programmes aimed to encourage safer behaviour across the workforce.

Programmes may include periodic modules, phishing examples, password guidance, reporting procedures and simulated phishing exercises. Some also offer active practice, reminders or behavioural evaluation. The medium or frequency by itself does not establish whether the training transfers to ordinary work [2, 3].

The intention behind these initiatives is clear and well-founded.

The human element appears in a substantial proportion of recorded breaches. Verizon’s 2026 Data Breach Investigations Report discusses human involvement under defined incident categories; that involvement must not automatically be described as employee negligence or a single root cause. Training is one possible protective measure within a wider system of technical and organisational controls [14, 15].

In principle, the logic appears sound. If employees understand how cyber attacks work, recognise the warning signs of malicious activity, and follow established security policies, the likelihood of successful attacks should decrease.

Awareness programmes can communicate an organisation's expectations and establish useful shared knowledge. Their contribution becomes clearer when it is assessed against specific outcomes, from what participants understand to what they can do in practice.

For an organisation, awareness training can be a practical way to communicate expectations, record participation and introduce shared security language. These functions have value, but completion alone does not guarantee compliance with every applicable obligation or demonstrate independent capability [3].

Evidence reviews show why knowledge and behaviour should be considered separately. A meta-analysis of 69 studies reported a positive pooled effect of training overall, while the estimated effect on observed behaviour was smaller and uncertain [16]. Knowledge gains and practical action therefore need distinct measures.

If employees already understand what cyber threats look like, why do attacks that rely on human interaction continue to succeed?

2. The Reality: Breaches Continue Despite Training

Despite the widespread adoption of cybersecurity awareness programmes, cyber incidents involving human behaviour remain extremely common.

Organisations use combinations of online learning, periodic exercises, simulated phishing and policy acknowledgement. Their stated purposes range from communicating expectations to developing practical action. Whether those activities achieve their aims must be assessed against a relevant outcome, not inferred from their presence [3].

To understand where training helps, we need evaluations that observe defined actions under specified conditions. Controlled and longitudinal studies give a clearer view of those effects than incident counts alone [4, 5].

The UK Government’s Cyber Security Breaches Survey 2025/2026 found that 43% of businesses and 28% of charities identified at least one cyber security breach or attack in the preceding 12 months. Among businesses identifying an event, 88% reported phishing. These categories include attempted attacks and do not describe a uniformly trained group [1].

These figures show the scale of reported exposure. Understanding the contribution of training requires a different question: how did the people, processes and safeguards involved affect the outcome?

The 2026 Verizon Data Breach Investigations Report identified a human element in 62% of recorded breaches under its definitions. Human involvement can coexist with credential compromise, technical exploitation and weaknesses in systems or processes; it is not a synonym for employee error [14].

Training history and knowledge should not be presumed from a breach report. What field evaluations do show is that trained participants can still respond to simulated phishing, and that the observed outcome depends on the intervention, environment and measure used [4, 5].

> Yet breaches still occur.

Major UK incidents also illustrate the consequences of disruption and the need to examine the full chain of events, from initial access through to containment and recovery.

In 2025, Marks & Spencer disclosed the theft of customer data following a cyber incident. The disclosure establishes that reported impact without identifying a verified employee or training-related cause [17].

The 2023 ransomware incident affecting Knights of Old, a transport business, was followed by the company's collapse. Subsequent reporting described a compromised-password account of the attack; that account should be understood as attributed reporting rather than an independent forensic reconstruction [18].

These incidents show why organisational resilience matters. Examining how an attack develops also means looking at identity controls, approval processes, monitoring and recovery arrangements rather than stopping at one visible action.

Phishing and social engineering may use apparent authority, urgency or familiar operational context to encourage action. The effects of particular cues depend on the situation and the recipient; they are not fixed or interchangeable across all attacks [19, 10].

Under these conditions, employees may make decisions that contradict what they learned during cybersecurity training.

The final action alone cannot tell us whether someone was careless. Time pressure, workload, instructions, accessible controls and the organisation's response to delay can all shape a decision. Human-factors analysis asks what made the action reasonable or likely in that setting [15, 7].

As a result, employees who understand cybersecurity risks may still make unsafe decisions in the moment.

A central limitation of information-only delivery is that learning about risk does not automatically produce a secure action when the person is working under competing demands. The relevant question is whether the training prepares the required action and whether the working environment makes it feasible [11, 9].

To understand why this gap exists, it is necessary to examine how employees actually make decisions within their daily work environments.

3. The Core Problem: Knowledge Alone Is Not Enough

When a course explains a security procedure, it can be tempting to assume that knowing the rule will be enough. In practice, the secure response may also depend on access to a tool, a manager's permission to delay and a workable route when the usual process fails [3].

Education can supply knowledge and clarify what is expected. It cannot, by itself, ensure that the necessary tools are available, that a manager permits a delay or that a procedure can be followed in the time and conditions of the task.

However, the reality of cybersecurity decision-making is far more complex.

Knowing the appropriate action does not mean it will always occur. Decisions depend on capability, opportunity and motivation, including the competing demands surrounding a task.

Training environments and live work may differ, although well-designed field simulations can capture some of those real conditions [9, 4].

Cyber risks can appear during ordinary work, rather than in clearly labelled security exercises. An email may arrive while someone is handling several tasks; a request may resemble a legitimate instruction from a colleague or supplier. Research into phishing difficulty and multitasking supports examining the context in which such messages are received [20, 8].

Some situations require a quick operational decision; others should permit time for verification. The important organisational question is whether staff have both an appropriate control and the authority to use it.

A fraudulent request can appear designed to fit the task its recipient is already trying to complete.

Modern social engineering attacks are specifically designed to exploit human cognitive shortcuts and workplace pressures. Rather than relying solely on technical exploitation, attackers manipulate urgency, authority, and trust in order to influence behaviour.

A phishing message may claim that a payment must be processed immediately to avoid disruption to a supplier. A fraudulent email may appear to come from a senior executive requesting confidential information. A malicious link may be embedded within what appears to be a routine system notification.

In each case, the attack attempts to trigger a rapid behavioural response before the recipient has time to question the legitimacy of the request.

People use familiar patterns and judgement shortcuts to work efficiently. Those shortcuts are often useful, but they can mislead when cues are unreliable or deliberately manipulated. Workload can alter attention, while the quality of a decision also depends on experience and feedback [21, 8].

Operational context further amplifies this challenge.

In some workplaces, efficiency and responsiveness are strongly rewarded. Someone may reasonably anticipate criticism for delaying a supplier payment or interrupting a senior colleague. When verification adds effort or lacks an approved alternative, a shortcut may seem like the easiest way to complete legitimate work [7, 22].

As a result, individuals may bypass verification steps or act on incomplete information in order to maintain workflow momentum.

A workaround can be understandable within competing organisational priorities without being safe. The useful response is to identify both the capability the person needed and any process, management or technology constraint that made the approved action difficult [22, 15].

Sometimes a person lacks the required knowledge; sometimes they know the guidance but cannot apply it. Correctly answering a training question does not settle which explanation fits a later workplace decision.

The challenge arises when individuals must apply that knowledge in complex, ambiguous situations where multiple behavioural pressures are operating simultaneously.

Knowledge alone does not guarantee the right action in these moments. Capability, motivation and a workable opportunity are also needed [9].

People need to recognise relevant uncertainty, use appropriate verification, complete secure tasks and escalate concerns within their role. Training can develop these abilities, but it cannot replace missing contact records, unavailable approvals or conflicting management instructions [3, 10].

An information-only course may leave too little opportunity to practise those actions. Yet contemporary awareness programmes differ considerably, and some already include realistic exercises, feedback and behavioural assessment. The useful distinction is between actual learning ingredients and outcomes, not two labels treated as mutually exclusive [2, 3].

It is therefore possible to improve understanding without a commensurate improvement in measurable conduct. That possibility should be tested with suitably designed performance measures rather than assumed from either completion rates or continuing attacks [16].

This gap between knowledge and behaviour represents one of the most important challenges in modern cybersecurity strategy.

Where the objective is better real-world decisions, learning programmes should identify the action required, create opportunities to practise it, assess performance and examine the conditions that may obstruct it [11, 3].

4. The Gap in Current Cybersecurity Training

One important question about cybersecurity learning is how success is measured. A useful evaluation needs to distinguish training delivery, knowledge, independent action and later workplace behaviour.

Completion records, quizzes and policy acknowledgement can supply useful information. They show that learning was delivered, that someone participated or that sampled knowledge was demonstrated. They cannot by themselves confirm that a person will carry out a consequential verification task during live work [11, 3].

These metrics are useful for demonstrating that training has been delivered. However, they reveal very little about how employees actually behave when confronted with a real cyber threat.

Phishing simulations can measure specific actions, such as clicking, submitting information or using a reporting button. Interpretation depends on how difficult the message was, who received it and which technical systems recorded the action. A simulation result is not automatically a measure of overall competence [20].

It is also inaccurate to assume that simulated phishing always happens in a quiet classroom. Unannounced field simulations can arrive during ordinary work among legitimate emails; announced exercises, field simulations and genuine attacks should be assessed as different settings [4, 5].

The question is how well the task and measurement capture the conditions relevant to the capability being assessed. Realistic context helps, but it does not make an exercise identical to a live incident.

Some longitudinal evaluations have observed initial improvement followed by weaker performance, and have examined how reminders may help. Their findings make delayed assessment useful, particularly when message difficulty and exposure vary over time [23].

Similarly, high completion rates for awareness training modules can create the impression that the workforce is well prepared to handle cyber threats. In reality, these programmes often rely on passive learning formats such as slide-based presentations or short video courses, followed by simple multiple-choice assessments.

Employees may successfully answer questions about phishing indicators or password policies, yet still struggle to apply this knowledge in real operational situations. Knowing that suspicious emails often contain unusual requests does not necessarily help an employee evaluate a convincing message that appears to come from a trusted colleague.

Incident reporting offers another example. People may hesitate because they are uncertain, doubt that a report will help or anticipate an unhelpful response. Research suggests that these barriers can influence reporting, while their importance depends on the organisation and its processes [24, 25].

An accessible reporting route, constructive response and capacity to triage concerns matter alongside a person's confidence. Teaching someone to report will not solve a mailbox that is unmonitored or a manager who discourages escalation [25, 26].

The result is a paradox within many organisations. Security training metrics may indicate that employees are well informed about cyber risks, while operational evidence suggests that unsafe decisions and missed warning signs continue to occur.

In other words, organisations are often measuring knowledge rather than behaviour.

Different measures answer different questions. Completion shows participation; a quiz can demonstrate sampled knowledge; guided practice shows what someone can do with help; and an unfamiliar, uncoached task tests independent capability. Workplace transfer needs later evidence from ordinary work, with measures such as phishing click rates interpreted in light of task difficulty [11, 20].

Measuring capability and diagnosing organisational conditions together gives a more useful account of the remaining gap. Measurement alone, however, does not close it; the organisation still has to make the appropriate action workable.

Recognising this gap is the first step toward developing a more effective approach to cybersecurity training.

5. The Shift Toward Behaviour-Led Cybersecurity Training

The practical case for behaviour-led learning sits within an established human-centred security tradition. Usable-security research has long examined the burden that controls place on people, while current institutional guidance already asks organisations to define behaviours and evaluate learning. The task is to put those principles into effective practice, not to portray them as newly discovered [27, 3].

Cybersecurity decisions are shaped by people, technology and organisational arrangements together. Firewalls, endpoint protection and identity controls can prevent or contain many failures; appropriate human action can also interrupt risk or support response. A complete account does not make either layer solely responsible [12, 10].

Social engineering remains an important attack route where a request mimics familiar work or apparent authority. Such attempts can also involve compromised credentials, technical exploitation or gaps in verification and approval [10, 14].

AI tools can help attackers generate convincing messages, impersonations and supporting material, and can increase the scale of some established techniques. These are assessed capabilities rather than proof that every resulting message is convincing or every attack succeeds [28].

A fraudulent email may closely resemble an organisation's usual language, while a fake website may reproduce familiar login features. Attackers can also use publicly available information to make a request seem plausible. None of these mechanisms requires flawless imitation; appearance alone is an unreliable basis for authorising a consequential action [28, 10].

Warning signs such as unusual domains or poor formatting can still be useful, but they cannot be the sole basis for judging a request. Where an instruction changes payment details, grants access or exposes sensitive information, the relevant approval and independent verification matter even when the message looks convincing [10, 28].

Risk recognition and verification should therefore work together. Some high-impact transactions require an approved check whether or not the message looks suspicious.

Behaviour-led learning responds by strengthening the decisions and practical actions people need within ordinary workflows, rather than relying exclusively on spotting suspicious-looking messages.

Rather than relying only on lists of warning signs, the proposed training design encourages people to examine what a request would authorise, select a trustworthy verification route and escalate concerns when needed. Whether those actions become reliable habits in work must be measured [29, 11].

This approach is consistent with human-centred security design: systems, processes and learning should account for realistic workload, predictable error and the demands of legitimate work. The underlying research is established; this paper proposes one practical way to organise the capability component [27, 15].

The capability targets are concrete: recognise a consequential risk, complete the appropriate check and raise uncertainty through a reliable route. The design must also make visible the tools and organisational support needed to carry out each action.

For an action to become repeatable, people need suitable opportunities to practise it, useful feedback and conditions that allow them to use it. Habit and transfer research support examining that combination over time [30, 11].

A capability framework can help organise this work around the decisions people need to make, rather than around a list of threats alone.

6. The Cyber Rebels Five-Domain Model

Building on the analysis of everyday working decisions in *Where Awareness Fails: Why Cybersecurity Training Isn't Stopping Breaches* [6], the Five-Domain Model brings recognition, verification, secure action, escalation and professional judgement into one practical structure. It sits alongside technical controls and organisational responsibility because the conditions of work affect whether secure action is possible [9, 12].

The Cyber Rebels Five-Domain Model is a practitioner-developed framework informed by behavioural science, human-factors research and evidence on learning. Its contribution is to organise five connected capabilities around everyday decisions so that training and assessment can focus on the action required.

Each domain poses a different question: did the person recognise the decision, verify what mattered, use a secure route, report appropriately and exercise judgement within the responsibilities of their role?

This approach builds on, rather than replaces, established research and guidance. NIST SP 800-50r1 addresses learning design, behaviour and evaluation; NIST CSF 2.0 places workforce learning within wider risk management; and human-factors research explains why systems and working conditions matter alongside individual capability [3, 12, 15].

The model deliberately avoids treating employees as the "weakest link". Individual capability matters, but safer action also relies on processes, permissions, resources, leadership and technical protections. Problems created by those conditions should not be scored as personal failings [15, 7].

Contextual Risk Recognition

The first domain focuses on an employee's ability to recognise when a situation may involve cyber risk.

Phishing messages and fraudulent requests can resemble normal business communication. Contextual Risk Recognition therefore concerns recognising when a consequential action requires scrutiny, even if the message looks familiar. Some transactions require a check by default, without first identifying a suspicious cue [20, 10].

Contextual Risk Recognition therefore focuses on developing the ability to interpret the surrounding circumstances of a request rather than relying solely on obvious technical indicators.

Consider a finance employee who receives an urgent request to change supplier payment details. The email carries familiar branding and arrives during a busy approval period. Looking for spelling mistakes may help, but it does not settle whether the change is authorised. The important question is what the request would change, which process applies and whether an independent verification is required.

- Does the request require an independent check under the organisation's financial procedure?
- Where can the employee obtain a trusted supplier contact?
- Who can authorise the change if the usual approver is unavailable?

Recognising the uncertainty gives the employee a reason to use the appropriate procedure. Whether they can complete it depends on accessible contact details, approval cover and permission to delay the transaction.

The training aim is to recognise when a consequential request requires a check, even if the message appears routine, and to choose a proportionate next step. The organisation must then make that approved step possible.

Verification and Control Discipline

Verification is not always a response to suspicion. Some consequential requests require a check as part of the normal process, however familiar or convincing they appear.

Verification and Control Discipline focuses on completing appropriate checks before consequential action. Some social-engineering attempts use urgency and trusted relationships to encourage a shortcut, but the organisation also has to provide a procedure people can follow [10, 7].

For example, a business email compromise attempt may impersonate a senior executive or trusted supplier and request a payment or sensitive information. Acting without the required authorisation can expose the organisation to loss, although other controls may still interrupt the attempt.

Verification may involve a call to a number obtained from independently maintained records, a separate approved system or a defined financial authorisation process. Merely switching channels is not enough if the contact information still comes from the request being checked [10].

A usable verification process can interrupt some attack pathways by establishing identity or authority before action. How consistently staff use that process, and what effect it has on outcomes, depends on both capability and organisational conditions.

Secure Operational Behaviour

The third domain focuses on integrating secure behaviour into daily working practices.

Policies may define how to handle data, credentials, systems and devices. Yet a person can know the approved route and still face delays, access problems or contradictory pressures while completing legitimate work. Those conditions need to be examined rather than treated as evidence of indifference [22, 31].

Secure Operational Behaviour focuses on embedding security practices into normal workflows so that they become routine rather than exceptional.

For example, employees may be encouraged to adopt secure habits such as locking devices when leaving desks, using password managers instead of reusing credentials, and storing sensitive information within approved systems rather than external platforms.

These behaviours may appear simple, but they play a significant role in reducing the attack surface available to cybercriminals.

Predictable processes can help people notice a departure from normal practice. But an attacker may imitate the expected process, so important controls should remain in place whether or not anything looks unusual.

Secure practices can limit particular opportunities for attack when the underlying systems and permissions support them. Training cannot substitute for secure defaults, available tools or leadership action to repair unusable processes [10, 27].

Incident Judgement and Escalation

Even in well-secured environments, employees will occasionally encounter situations that appear suspicious but unclear. In these moments, the ability to exercise judgement and escalate concerns appropriately becomes critical.

Incident Judgement and Escalation concerns recognising a possible problem, knowing the appropriate route and making a useful report without waiting for complete certainty. Confidence matters, but so do the reporting system and the response people receive [25].

Traditional training often encourages employees to report suspicious emails or unusual activity, but many staff hesitate to escalate concerns because they fear being wrong or wasting the security team's time.

Training can normalise reporting uncertainty and allow people to practise the route. It does not remove the barrier if that route is difficult to access, no one responds or reporting attracts criticism.

A person who receives a suspicious request for login credentials may be better served by using the approved reporting route than by investigating it personally. A useful report can help a security team identify related messages, provided the team can triage and respond [4, 10].

A useful report is only the beginning of an effective response. The organisation also needs to acknowledge it, assess what it means and act where necessary. Reporting exercises should therefore show what happens after someone raises a concern.

Professional Cyber Judgement

The fifth domain addresses professional judgement under competing responsibilities: deciding what action is proportionate, whose authority is required and how to act when service, privacy, safety and security priorities interact. Culture and responsibility matter, but they do not exhaust the capability.

Cybersecurity decisions arise throughout an organisation, not only within specialist teams. Staff have role-appropriate responsibilities when handling information and requests, while leaders and technical owners remain accountable for governance, usable controls and accepted risk [12, 3].

Professional Cyber Judgement encourages employees to view cybersecurity decisions as part of their role rather than as external rules imposed by security teams.

In practice, this means fostering a culture where individuals feel responsible for questioning unusual requests, protecting sensitive information, and supporting organisational security processes.

For example, an employee who receives an unusual request from a colleague may feel comfortable asking clarifying questions rather than assuming the request must be legitimate. Similarly, staff may proactively report suspicious communications rather than ignoring them.

Professional judgement depends on more than willingness to challenge a request. People need clear decision rights, a realistic escalation route and leadership support when security requirements appear to conflict with legitimate work.

Building Behavioural Cyber Resilience

Together, the domains give a practical structure for training and assessment, from recognising and checking a request to using secure procedures, reporting concerns and making a defensible decision.

Within layered security, these capabilities can give staff a clearer contribution to prevention and response when the organisation provides workable processes, appropriate authority and reliable tools.

Model scope and future evaluation

The Five-Domain Model draws on established research, while its particular arrangement and the effectiveness of training built around it still require direct evaluation. The first questions are whether the domains cover relevant decisions and whether assessors can judge performance consistently. Later studies can test whether a defined programme improves independent decisions in work and supports better operational outcomes.

A practical sequence would begin with role-relevant task analysis, develop scoring criteria for unfamiliar scenarios, compare the resulting learning approach with a credible alternative and examine later performance where the approved action was genuinely available. Knowledge, guided exercise performance, independent capability, workplace transfer and incident outcomes must remain separate levels of evidence [11, 3].

7. What Behaviour-Led Training Looks Like in Practice

An information-led activity may primarily explain threats or procedures. A capability-focused activity adds opportunities to apply the relevant judgement and assess the action. Some existing awareness programmes already include these features, so the comparison should be based on methods rather than labels [2].

The design difference is concrete: specify the decision, practise the action, give task-relevant feedback and distinguish guided discussion from unprompted performance.

Slides, videos, examples and multiple-choice questions can help people acquire knowledge. They are insufficient on their own when an organisation wants to know whether someone can carry out a demanding task independently. The assessment should match the intended learning outcome [3].

The proposed behaviour-led approach combines knowledge with role-relevant practice: recognising uncertainty, following an approved route, explaining why it fits and responding when the ordinary route fails. Research on practice and transfer supports assessing these actions after learning and at later opportunities to use them [32, 11].

Participants are placed in scenarios that mirror the types of decisions they encounter in their daily work. Instead of being told the correct answer in advance, they are asked to analyse situations, question assumptions, and determine how they would respond.

For example, a training exercise may present a realistic email from what appears to be a senior colleague requesting urgent action. The message may contain no obvious indicators of fraud. Participants are asked to decide whether they would act on the request, what questions they would ask, and how they would verify the instruction.

Through guided discussion, participants explore the reasoning behind their decisions and identify the behavioural cues that indicate potential risk.

Discussion can make assumptions visible. Independent capability becomes clearer when participants carry out an unfamiliar task without prompting. Practice and feedback should give them opportunities to improve before that assessment [32, 33].

A practical teaching principle is "Pause and Verify". Social-engineering messages may press for immediate action through urgency, apparent authority or a familiar process. The pause creates an opportunity to establish whether the request requires independent verification or further authorisation.

A pause creates space to consider what an action would authorise and what must be confirmed. It is useful only if there is a reliable procedure to follow, including an exception route when a contact or approver is unavailable. High-impact routine requests may need verification even without an unusual warning sign.

The check might use independently maintained contact details, a required approval in an authorised system or an agreed escalation route. An attacker-provided telephone number is not an independent source. Practice should assess the completed check, not merely whether someone says they would pause [10].

Behaviour-led training therefore shifts the emphasis from recognising obvious threats to managing uncertainty.

Realistic operational context matters because messages and requests arrive among legitimate tasks, interruptions and competing priorities. Some attacks exploit those conditions, but not every case requires the same level of pressure. Scenarios should reproduce the demands that matter for the decision being practised [8, 20].

Training exercises therefore reflect these conditions. Participants may analyse communication chains, evaluate conflicting information, or discuss how organisational pressures such as urgency or hierarchy influence decision-making.

Examining these decisions openly helps participants see why a shortcut may feel reasonable when time, authority or the demands of the task make an immediate response attractive.

AI-assisted communications can appear highly convincing, making visual cues alone an incomplete defence. Sound verification relies on what is being requested and a trustworthy source of authority, supported by technical measures that limit exposure [28, 10].

The proposed training can also cover when and how to raise a concern. Its immediate educational aim is that participants recognise an escalation threshold and can use the organisation’s reporting route; improved live response is a further outcome to test.

Reporting exercises should include uncertainty, the possibility that a mistake has already been made and the response participants can expect. Normalising escalation may be useful, but earlier detection depends on accessible routes, constructive management and effective triage [25, 26].

Importantly, behaviour-led training is designed to be interactive and discussion-driven rather than passive. Participants are encouraged to share experiences, challenge assumptions, and explore how cybersecurity risks intersect with their specific roles.

This collaborative approach reinforces the idea that cybersecurity is not solely the responsibility of technical teams, but a shared organisational capability.

Practice and constructive feedback can help people become more comfortable questioning a request or reporting uncertainty. Later tasks and observations in work can show whether those responses are becoming more reliable.

Published training evaluations show mixed results. Some phishing interventions have improved measured outcomes, while large workplace studies found limited or absent benefit from the particular embedded training they tested. This makes careful design and follow-up assessment essential [34, 4, 5].

Sustained secure practice depends on repeated opportunities, accessible controls and feedback. Delayed observation, rather than a classroom result alone, is needed to understand whether a response becomes routine [30, 11].

The aim is to move from understanding, through practice, towards independent action that people can use in their working environment.

8. Implications for Organisations

Where decisions made during ordinary work contribute to cyber exposure, organisations need to support practical capability as well as technical and governance measures. The balance depends on the actual incident pathways and constraints, not on a presumption that human behaviour is always the primary cause.

Technology remains essential: firewalls, endpoint protection, identity controls, monitoring and secure configuration can prevent incidents or reduce their impact. Learning should be designed to work with these defences, not to compensate for their absence [12, 10].

While these technologies are essential, they address only part of the problem.

Some attacks begin with apparently ordinary work: a payment instruction, an email, a call or a routine system action. Others begin through technical exploitation. A single incident may involve both, with several opportunities for prevention or containment [14].

Cybersecurity decisions and technical infrastructure are therefore interdependent. A human-first approach does not mean treating technology as secondary; it means making the conditions of real work part of prevention and response.

This reality has significant implications for how organisations design their cybersecurity strategies.

Cybersecurity Is Not Just an IT Responsibility

If employees across the organisation interact with systems, data, and communications, then cybersecurity cannot be treated solely as a technical responsibility.

Finance teams may receive fraudulent payment requests. Human resources departments handle sensitive personal information that attackers seek to obtain. Customer service staff communicate with external parties who may attempt impersonation attacks. Senior leaders may be targeted through highly tailored social engineering campaigns.

Everyday decisions are one opportunity for prevention, alongside identity controls, authorisation, monitoring and recovery arrangements.

To understand this challenge more clearly, it is helpful to recognise that modern cybersecurity operates across three distinct layers of organisational responsibility.

Technical teams remain responsible for implementing and maintaining technical security controls such as firewalls, endpoint protection, identity management systems, and network monitoring tools. These controls form the technological foundation of the organisation's security architecture.

Leadership, governance and relevant support functions establish policy, acceptable risk, workable routes, training expectations and the resources required to maintain them. These responsibilities cannot be transferred wholesale to employees.

Human decisions influence how controls are used and when concerns are raised, while technical design, defaults and governance also determine whether those decisions are necessary, feasible and effective.

This introduces a third and often overlooked layer of defence: behavioural controls.

The behavioural contribution appears when people verify a consequential request, handle information through an approved route or report a possible incident. Whether they can do those things depends on the tools, authority and organisational signals they encounter [7].

In practical terms, cybersecurity can therefore be understood as a shared organisational capability:

Technology and security functions maintain technical protections, identity controls and monitoring.

Leadership and governance define responsibilities, workable processes, risk decisions and support.

The wider workforce uses the available controls, makes role-appropriate decisions and reports concerns.

People can provide a valuable additional opportunity to interrupt risk, but protective systems should also prevent or limit the consequences of ordinary mistakes. Better verification and escalation are shared capabilities requiring organisational enablement, not simply more vigilance [15, 10].

Cybersecurity is therefore an organisational capability: staff, managers and specialists each play different roles in decisions, prevention and response. The organisation remains responsible for making secure action realistic.

Training programmes, therefore, must be designed not only for IT professionals but for the wider workforce whose everyday decisions influence cybersecurity outcomes.

Moving Beyond Compliance-Based Training

Another implication concerns how organisations approach cybersecurity education.

Many awareness programmes are delivered primarily to satisfy regulatory or compliance requirements. Employees complete training modules, pass knowledge-based assessments, and acknowledge organisational policies.

While these activities demonstrate that training has been delivered, they do not necessarily ensure that employees are capable of responding securely in real-world situations.

If the objective is to reduce cyber risk, training must focus on developing behavioural capabilities such as recognising unusual requests, verifying instructions, questioning unexpected activity, and escalating potential incidents.

This requires training that reflects real operational conditions rather than simply delivering theoretical knowledge.

Measuring Behaviour Rather Than Training Completion

Organisations must also reconsider how they measure the effectiveness of cybersecurity initiatives.

Completion rates and assessment scores provide limited insight into how employees behave when confronted with real threats. An employee may correctly identify phishing indicators during a training quiz but still respond to a convincing social engineering message during a busy workday.

Useful measures depend on the capability: adequate verification among transactions requiring a check; timely, actionable reports relative to reporting opportunities; or safe completion of role-appropriate tasks. Counts alone can mislead when exposure, task difficulty or reporting capacity changes.

These measures can illuminate workforce contributions if they are interpreted alongside technical performance, organisational conditions and appropriate denominators. None should be treated as an automatic measure of incident reduction.

For example, an increased number of reports can reflect greater willingness to raise concerns, more exposure or more duplicate submissions. A verification rate needs a denominator of transactions that required verification. Fair evaluation must consider whether the organisational route was available and whether the action was appropriate.

Embedding Security Into Everyday Work

For behaviour-led cybersecurity to succeed, secure behaviour must be embedded within everyday operational processes.

People are more likely to have a usable secure route when procedures fit their real tasks. Accessible verification contacts, clear escalation channels and sufficient authority help. Some justified security checks will delay work; the aim is proportionate friction, not an impossible promise of zero interruption [7, 22].

When security behaviours are integrated into normal workflows, they become part of routine professional practice rather than additional tasks imposed by technical teams.

The Workforce as a Security Control

Employees are not merely sources of risk. They can contribute to prevention, reporting and recovery through the decisions they make within their roles, supported by effective technical controls and management processes.

Employee reporting can complement technical monitoring by bringing attention to suspicious activity encountered during work. Its value depends on report quality, response capacity and how quickly the organisation can act [4].

With effective systems and organisational support, the workforce can contribute to layered defence through sound decisions, appropriate verification and timely reporting.

Organisations should test these capabilities against what happens during real work, including obstacles that training alone cannot remove.

Conclusion

Cybersecurity awareness training has long been used to explain phishing, credential security and safer online conduct. It can communicate useful knowledge and organisational expectations, although the methods and evidence of effectiveness vary substantially [2].

The evidence points to a practical distinction: people can learn about a threat without reliably applying the required response. Training transfer also depends on the workplace in which that response is needed [16, 11].

Cyber risks may arise in busy, familiar working conditions where a person is trying to complete a legitimate task. Pressure, authority cues and uncertainty can shape the next action; so can the availability of a reliable verification route and permission to use it [8, 7].

Some failures involve missing knowledge. Others arise when a person understands the risk but lacks a workable opportunity, appropriate authority or practised response. The appropriate intervention depends on diagnosing which conditions shaped the decision [9].

> This distinction is critical.

A programme that measures only knowledge cannot establish whether people can act appropriately under uncertainty. Capability assessment and organisational diagnosis are needed to understand that remaining question.

Reducing human-related cyber risk requires a different approach – one that recognises cybersecurity as a behavioural capability embedded within everyday organisational activity.

Behaviour-led training is one practical contribution to that approach, provided its methods and outcomes are specified and evaluated.

The proposed training design targets contextual risk recognition, verification, secure operational behaviour, incident judgement and professional cyber judgement. It aims to develop and assess actions, rather than assuming that exposure to information produces them.

The Cyber Rebels Five-Domain Model draws on established behavioural, human-factors and learning research. It brings those foundations together into a practical framework for recognising risk, verifying what matters, acting securely, escalating uncertainty and exercising professional judgement.

AI-assisted deception can make some requests more convincing and familiar in appearance. That reinforces the importance of clear authorisation, appropriate verification and usable controls rather than reliance on visible errors alone [28, 10].

Technical protections and governance remain essential. Human decisions matter within those layers, but a secure outcome also depends on design, resources, authority and response arrangements.

The stronger case is therefore not for awareness to disappear, but for learning to extend beyond it: explain the risk, practise the decision, support the approved action and evaluate whether it transfers into work.

The next step is direct evaluation: assessing the capabilities reliably and examining whether training helps people apply them in their roles over time.

References

[1] Department for Science, Innovation and Technology and Home Office (2026) Cyber Security Breaches Survey 2025/2026. GOV.UK. https://www.gov.uk/government/statistics/cyber-security-breaches-survey-20252026/cyber-security-breaches-survey-20252026

[2] Prümmer, J., van Steen, T. and van den Berg, B. (2024) 'A systematic review of current cybersecurity training methods', Computers & Security, 136, 103585. https://doi.org/10.1016/j.cose.2023.103585

[3] Merritt, M., Hansche, S., Ellis, B., Sanchez-Cherry, K., Nethery Snyder, J. and Walden, D. (2024) Building a Cybersecurity and Privacy Learning Program. NIST SP 800-50r1. Gaithersburg, MD: NIST. https://doi.org/10.6028/NIST.SP.800-50r1

[4] Lain, D., Kostiainen, K. and Čapkun, S. (2022) 'Phishing in organizations: findings from a large-scale and long-term study', IEEE Symposium on Security and Privacy, pp. 842–859. https://doi.org/10.1109/SP46214.2022.9833766

[5] Liu, L., Ho, G., Mirian, A., Luo, E., Tong, K., Lee, E., Longhurst, C.A., Dameff, C., Savage, S. and Voelker, G.M. (2025) 'Understanding the efficacy of phishing training in practice', IEEE Symposium on Security and Privacy. https://doi.org/10.1109/SP61157.2025.00076

[6] Longhurst, Andrew (2026) Where Awareness Fails: Why Cybersecurity Training Isn't Stopping Breaches. Cyber Rebels. https://doi.org/10.5281/zenodo.19360661

[7] Beautement, A., Sasse, M.A. and Wonham, M. (2008) 'The compliance budget: managing security behaviour in organisations', Proceedings of the New Security Paradigms Workshop, pp. 47–58. https://doi.org/10.1145/1595676.1595684

[8] Lu, X., Jiang, J., Head, M. and Yang, J. (2026) 'Phishing detection in multitasking contexts: the impact of working memory load, goal activation, and message framing cue on detection performance', European Journal of Information Systems, 35(1), pp. 134–164. https://doi.org/10.1080/0960085X.2025.2548543

[9] Michie, S., van Stralen, M.M. and West, R. (2011) 'The behaviour change wheel: a new method for characterising and designing behaviour change interventions', Implementation Science, 6, 42. https://doi.org/10.1186/1748-5908-6-42

[10] NCSC (n.d.a) Phishing attacks: defending your organisation. https://www.ncsc.gov.uk/pdfs/guidance/phishing.pdf

[11] Blume, B.D., Ford, J.K., Baldwin, T.T. and Huang, J.L. (2010) 'Transfer of training: a meta-analytic review', Journal of Management, 36(4), pp. 1065–1105. https://doi.org/10.1177/0149206309352880

[12] NIST (2024) The NIST Cybersecurity Framework (CSF) 2.0. NIST CSWP 29. Gaithersburg, MD: NIST. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf

[13] NCSC (2026) Cyber Essentials: Requirements for IT Infrastructure, version 3.3. April. https://www.ncsc.gov.uk/files/cyber-essentials-requirements-for-it-infrastructure-v3-3.pdf

[14] Verizon (2026) 2026 Data Breach Investigations Report. https://www.verizon.com/business/resources/T1e0/reports/2026-dbir-data-breach-investigations-report.pdf

[15] Reason, J. (2000) 'Human error: models and management', BMJ, 320(7237), pp. 768–770. https://doi.org/10.1136/bmj.320.7237.768

[16] Prümmer, J., van Steen, T. and van den Berg, B. (2025) 'Assessing the effect of cybersecurity training on end-users: a meta-analysis', Computers & Security, 150, 104206. https://doi.org/10.1016/j.cose.2024.104206

[17] Marks and Spencer Group plc (2025) 'Cyber incident further update'. Regulatory announcement, 13 May. https://data.fca.org.uk/artefacts/NSM/RNS/5651766.html

[18] Millett, C. (2025) 'Employee's weak password opened door to Knights of Old's cyber attackers, BBC documentary reveals', Motor Transport, 21 July. https://motortransport.co.uk/operator-news/employees-weak-password-opened-door-to-knights-of-olds-cyber-attackers-bbc-documentary-reveals/26654.article

[19] Parsons, K., Butavicius, M., Delfabbro, P. and Lillie, M. (2019) 'Predicting susceptibility to social influence in phishing emails', International Journal of Human-Computer Studies, 128, pp. 17–26. https://doi.org/10.1016/j.ijhcs.2019.02.007

[20] Steves, M., Greene, K. and Theofanos, M. (2020) 'Categorizing human phishing difficulty: a Phish Scale', Journal of Cybersecurity, 6(1), tyaa009. https://doi.org/10.1093/cybsec/tyaa009

[21] Kahneman, D. and Klein, G. (2009) 'Conditions for intuitive expertise: a failure to disagree', American Psychologist, 64(6), pp. 515–526. https://doi.org/10.1037/a0016755

[22] Herley, C. (2009) 'So long, and no thanks for the externalities: the rational rejection of security advice by users', Proceedings of the New Security Paradigms Workshop. https://www.microsoft.com/en-us/research/publication/so-long-and-no-thanks-for-the-externalities-the-rational-rejection-of-security-advice-by-users/

[23] Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., von Landesberger, T. and Volkamer, M. (2020) 'An investigation of phishing awareness and education over time: when and how to best remind users', Proceedings of SOUPS 2020. https://www.usenix.org/conference/soups2020/presentation/reinheimer

[24] Kwak, Y., Lee, S., Damiano, A. and Vishwanath, A. (2020) 'Why do users not report spear phishing emails?', Telematics and Informatics, 48, 101343. https://doi.org/10.1016/j.tele.2020.101343

[25] NCSC (n.d.b) 'Principle 2: Build safety, trust and processes', Cyber security culture principles. https://www.ncsc.gov.uk/collection/cyber-security-culture-principles/principle-2

[26] Edmondson, A. (1999) 'Psychological safety and learning behavior in work teams', Administrative Science Quarterly, 44(2), pp. 350–383. https://doi.org/10.2307/2666999

[27] Adams, A. and Sasse, M.A. (1999) 'Users are not the enemy', Communications of the ACM, 42(12), pp. 40–46. https://doi.org/10.1145/322796.322806

[28] NCSC (2025) Impact of AI on cyber threat from now to 2027. 7 May. https://www.ncsc.gov.uk/sites/default/files/pdfs/publication/impact-ai-cyber-threat-now-2027.pdf

[29] Gollwitzer, P.M. and Sheeran, P. (2006) 'Implementation intentions and goal achievement: a meta-analysis of effects and processes', Advances in Experimental Social Psychology, 38, pp. 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

[30] Lally, P., van Jaarsveld, C.H.M., Potts, H.W.W. and Wardle, J. (2010) 'How are habits formed: modelling habit formation in the real world', European Journal of Social Psychology, 40(6), pp. 998–1009. https://doi.org/10.1002/ejsp.674

[31] Stanton, B., Theofanos, M.F., Spickard Prettyman, S. and Furman, S. (2016) 'Security fatigue', IT Professional, 18(5), pp. 26–32. https://doi.org/10.1109/MITP.2016.84

[32] Morris, T.H. (2020) 'Experiential learning – a systematic review and revision of Kolb's model', Interactive Learning Environments, 28(8), pp. 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

[33] Pan, S.C. and Rickard, T.C. (2018) 'Transfer of test-enhanced learning: meta-analytic review and synthesis', Psychological Bulletin, 144(7), pp. 710–756. https://doi.org/10.1037/bul0000151

[34] Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L.F., Hong, J., Blair, M.A. and Pham, T. (2009) 'School of Phish: a real-world evaluation of anti-phishing training', Proceedings of SOUPS 2009, article 3, pp. 1–12. https://doi.org/10.1145/1572532.1572536