

Article

A Quantitative Evaluation of Cyber4Me: A Holistic Framework for Enhancing Individual Cybersecurity Awareness

Md. Arafatur Rahman ^{1,*}, Mohamad Ibrahim ^{2,*} , Bashir Ahmed ¹, Nadia Refat ¹, Tan Sze Wei ¹ and Prashant Pillai ¹

¹ School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton WV1 1LY, UK; b.ahmed19@wlv.ac.uk (B.A.); atoshi.refat@gmail.com (N.R.); s.w.tan@wlv.ac.uk (T.S.W.); p.pillai@wlv.ac.uk (P.P.)

² School of Computing Engineering and the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, UK

* Correspondence: arafatur.rahman@wlv.ac.uk (M.A.R.); mohamad.ibrahim@napier.ac.uk (M.I.)

Abstract

Human factors remain the dominant contributor to cybersecurity incidents, yet awareness training produces only moderate and often non-durable behaviour change, and most evaluated programs are either purely digital or evaluated only at the framework level. This study addresses two gaps: the scarcity of empirical and demographically stratified evidence for multi-modal community-facing awareness programs, and the lack of an explicit account of how artificial intelligence (AI) should be integrated into such programs rather than treated as an optional add-on. We evaluate Cyber4Me, a four-stage individual-awareness intervention (community roadshows, structured training, a hackathon, and a physical-digital escape room) that is wrapped in a cross-cutting AI adaptive layer built entirely on structured performance and behaviour data baseline competency tiering, awareness-behaviour gap detection, predictive early-warning, and personalised recommendation, with no reliance on free text. Using a single-group pre-post design with 130 participants in the UK Black Country region and a multi-dimensional Likert instrument, all four competency domains (confidence, familiarity, GDPR knowledge, incident-response preparedness) improved significantly (paired-*t*, all $p < 0.001$; large within-participant effects, Cohen's $d \geq 1.0$). Improvement was strongly moderated by demographics: older adults gained most in familiarity, undergraduates in confidence, and lower-education participants in regulatory knowledge. The contributions are as follows: transparent and demographically stratified pre-post evidence for a multi-modal awareness program with effect sizes reported; a fitness-for-purpose comparison against contemporary analogs (KnowBe4, Proofpoint, CyberPatriot, iCAT, CAT-RWE, GPT-CSAT, escape-room studies) that treats AI as a first-class design dimension; and an articulated AI integration architecture for the framework, demonstrated offline on the cohort using only structured performance and behaviour data (no free text). In this architecture, a gradient-boosted classifier assigns participants to three baseline competency tiers at 93.1% cross-validated accuracy; these tiers differ sharply in measured improvement (ANOVA $F = 68.8$, $p < 0.001$; Foundational +1.79 vs. Applied +0.30 scale points), an awareness-behaviour gap segment is detected and predicted from intake signals alone (AUC = 0.73), and a recommender routes participants to personalised follow-on tracks. As the design is single-group and self-reported, results are reported as evidence of within-participant change associated with the intervention rather than as a causal efficacy estimate, and the AI layer is demonstrated for feasibility rather than being evaluated as a separate trial arm; the scope is explicitly individual security awareness and behaviour, not technical network, IIoT, or cloud security.



Academic Editor: Paolo Bellavista

Received: 31 May 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 29 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: cybersecurity awareness; AI-adaptive training; demographic stratification; education and age moderators; multi-modal learning; behavioural change; human factors

1. Introduction

Cyber threats such as phishing, ransomware, supply chain compromises, and data breaches remain a major source of financial, operational, and reputational harm for individuals, organisations, and wider society. Global cybercrime damages have been projected to reach USD 10.5 trillion annually by 2025 [1] and human factors continue to play a central role in cybersecurity incidents, with a frequently cited estimate suggesting that 95% of breaches involve some form of human contribution [2]. Despite the growing investment in cybersecurity awareness and training, only 27% of organisations report confidence in their employees' security practices [3,4]. This persistent weakness highlights a central problem in cybersecurity education: while training initiatives are widely deployed, their ability to produce durable behavioural change remains limited. Recent peer-reviewed evidence supports this concern. The meta-analysis by Prümmer, van Steen, and van den Berg [5] shows that cybersecurity training can improve knowledge and threat awareness, yet its effect on long-term behavioural change is only moderate. This tension between the widespread use of awareness programs and their modest behavioural impact motivates the need for careful evaluation of alternative training designs, particularly those that account for participant characteristics such as age and educational background.

To avoid ambiguity flagged in earlier review, we state the scope precisely. This work concerns individual cybersecurity awareness and security behaviour, a human-factors problem for a general non-specialist community population. It is not an evaluation of technical security controls for enterprise networks, industrial IoT, or edge/fog/cloud infrastructure, which are distinct engineering domains evaluated with different instruments [6,7]. The competencies measured here (confidence, terminology familiarity, regulatory/GDPR knowledge, and incident response preparedness) are the behavioural antecedents that the human-factors literature links to real-world risk reduction [2,3,5].

Existing awareness programs have converged on digital, organisational, and increasingly AI-driven formats. KnowBe4 [8] and Proofpoint [9] use AI-scored phishing simulation at organisational scale; iCAT [10] couples knowledge graphs and serious games with adaptive micro-learning; CAT-RWE [11] tailors content for remote workers; and GPT-CSAT [12] uses generative models for per-trainee personalisation. A clear trajectory emerges from this body of work: the field is moving from static one-size-fits-all content toward AI-mediated adaptation, and recent evidence consistently treats adaptivity not breadth of content as the decisive lever for engagement and durable change [5,13]. By contrast, programs that reach broader public audiences through physical embodied learning (roadshows, hackathons, escape rooms) remain comparatively under-evaluated and have rarely been combined with an explicit AI layer [14–16].

This produces two specific under-addressed gaps. First, there is little transparent, demographically stratified pre–post evidence for multi-modal community-facing awareness programs; prior work tends to report either aggregate organisational metrics or framework-level feasibility, and rarely reports how age and education jointly moderate outcomes across several competency domains at once [5,17]. Second, AI is the single capability that most clearly differentiates state-of-the-art analogs; multi-modal physical programs have not articulated how AI integrates with embodied delivery, and AI tends to be deferred as future work rather than designed in. Therefore, the novelty of Cyber4Me lies not in any single mechanic—escape rooms, hackathons, and adaptive content already exist—but in

sequencing them as one progressive experiential program for a non-technical public and then wrapping that sequence inside a cross-cutting AI adaptive layer so that personalisation operates across physical stages rather than inside a solely digital module.

Accordingly, this study has three explicit objectives: first, to quantify within-participant change in four competency domains following the Cyber4Me intervention while transparently reporting the effect sizes; second, to characterise how age and education moderate both baseline competence and measured improvement; and third, to position Cyber4Me against contemporary analogs on a fitness-for-purpose basis, in which AI adaptivity is treated as a first-class dimension and the AI architecture is specified.

This paper makes the following contributions:

- An empirical and demographically stratified pre–post evaluation of Cyber4Me ($n = 130$), reporting effect sizes and subgroup moderation transparently;
- A fitness-for-purpose comparison against seven contemporary analogs (Table 1) that argues from suitability-to-audience rather than feature count and that treats AI adaptivity as the most consequential differentiating feature;
- An articulated AI adaptive-layer architecture (Section 3.7) built entirely on structured performance and behaviour data baseline competency tiering, awareness-behaviour gap detection, predictive early-warning, and follow-on recommendation demonstrated offline on the study cohort (Section 5.3). This includes a 93.1%-accurate tier classifier in which the tiers differ significantly in measured improvement, closing the design gap identified above;
- An explicit statement of the methodological boundaries of a single-group design, so that within-participant change is not overstated as causal efficacy.

The remainder of the paper is structured as follows: Section 2 reviews the literature on cybersecurity training, demographic moderators, gamified approaches, and analogous frameworks; Section 3 describes the Cyber4Me framework, its theoretical grounding, and its comparative positioning; Section 4 details the methodology, including recruitment, instrument design, and limitations; Section 5 reports the findings, while Section 6 discusses their implications for cybersecurity education and demographically responsive training design; finally, Section 7 concludes the paper and outlines directions for future work.

2. Literature Review

Cybersecurity awareness is not a one-size-fits-all endeavour; individual differences significantly influence how people perceive, learn, and apply security best practices. Research highlights that demographic factors such as age, education level, and gender play a crucial role in shaping vulnerability to cyber threats and the effectiveness of training interventions. Understanding these variations allows organisations to design tailored cybersecurity programmes that maximise engagement, knowledge retention, and behavioural change. This section reviews four threads: (i) demographic moderators of training effectiveness; (ii) pedagogical and gamification design principles; (iii) contemporary AI-enabled and remote worker training frameworks; and (iv) the use of cyber escape rooms and hackathons. Each thread closes with how the present study extends or differs from that work.

2.1. Demographic Moderators of Training Effectiveness

Age has a complex relationship with cybersecurity outcomes. Younger adults (18–29) report higher exposure to online threats but adapt readily to interactive training [18,19], while older adults often show slower processing of new technical concepts [20]. However, older workers possess what Salthouse [21] terms “crystallised intelligence” derived from accumulated work experience, which can support uptake of secure practices once content is presented in a familiar register. This duality has prompted calls for age-adaptive

training designs that combine simplified scenario-based material for older participants with more gamified content for younger users [22]. Recent work reinforces that older adults should be treated as a distinct design audience rather than a deficit group: Parti, Abdelhamid, and Ladancsik [23] analysed digital learning preferences for cybersecurity through a socioeconomic lens and found that older adults' engagement is shaped substantially by delivery format and social and economic context rather than by cognitive capability alone, supporting the use of embodied community-facing delivery of the kind that Cyber4Me adopts. Education has an analogous moderating role: postgraduate-educated participants typically display higher baseline self-efficacy and stronger phishing-avoidance behaviour, but training can substantially close the gap for participants with only secondary education [22,24–26]. Gender effects on training outcomes are generally smaller and less consistent than effects of age or education [27,28].

Prümmer et al.'s meta-analysis [5] synthesised the most recent end-user training evidence, reporting that training reliably improves knowledge and threat recognition but has a more moderate effect on durable behaviour change; method of delivery (interactive vs. presentation), social setting, and modernness of method were identified as significant moderators. The present study contributes to this literature by stratifying outcomes simultaneously by age and education across four competency domains, which the above meta-analysis identified as a remaining gap.

2.2. Pedagogical and Gamification Design Principles

Generic one-size-fits-all training has repeatedly been shown to underperform. Effective programs use clear and jargon-free language, culturally relevant examples, varied media [24], and continuous reinforcement with assessment [25]. A systematic review by Tatum [29] of gamified security-awareness programs reported engagement improvements of 40–60% relative to traditional lecture-based formats and identified that combining multiple gamification mechanics outperforms any single mechanic. The APPEARS framework of Birajdar et al. [14] (Awareness, Perception, Practice, Experience, Achievement, Reward, Socialisation) provides a structured way to assess which dimensions a given program addresses, and is used in Section 3.1 to position Cyber4Me's stages. Beyond pedagogy, the effectiveness of any training format is bounded by whether users accept and engage with it in the first place. Fallatah, Kävrestad, and Furnell [30] modelled user acceptance of cybersecurity training using the technology acceptance model; synthesising the prior literature, they identified regulatory control, worry, apathy, and trust as the factors that shape users' intention to engage with training. This acceptance perspective complements the gamification literature: voluntary, socially embedded, and gamified formats can lower apathy and build trust, which is part of the design rationale for Cyber4Me's community-facing team-based delivery. This is examined further through the lens of Self-Determination Theory in Section 3.1.

2.3. Contemporary Frameworks: AI-Enabled and Specialised Approaches

Several recent frameworks extend the older KnowBe4/Proofpoint paradigm; because earlier review observed that analogs were described only superficially, we characterise each by its underlying mechanism, evaluation evidence, target population, and the specific role that AI plays.

iCAT [10] represents the knowledge graph branch of adaptive training: it encodes the threat domain as a structured graph, drives serious game scenarios from that graph, and uses a micro-learning loop in which learner responses update a difficulty estimate that selects the next item. Its mechanism for personalisation is symbolic and rule-driven, and its principal evidence is a design-and-feasibility argument rather than a population-scale

pre-post measurement; this leaves open whether the adaptive loop changes behaviour at scale, as opposed to merely engagement.

CAT-RWE [11] targets the specific risk profile of remote workers (unmanaged networks, blurred work/home boundaries) and tailors content to role and context. Its AI contribution is content selection conditioned on a worker profile and it is validated through structured expert review rather than learner outcomes, so its claims concern construct coverage more than measured efficacy.

GPT-CSAT [12] sits on the generative branch; it uses large language models to synthesise per-trainee material conditioned on a profile, directly attacking the personalisation ceiling of fixed-content training. Its strength—open-ended generation—is also its main unverified risk, since generative output requires guardrails against inaccurate or unsafe security advice; like iCAT, it is reported at the framework level.

KnowBe4 and Proofpoint [8,9] dominate organisational deployment and report large multi-month reductions in simulated-phishing click rates. Their AI is mature but narrow; it scores and schedules phishing simulations and segments users by risk. While highly scalable, both are predominantly digital and compliance-oriented, with little embodied or community-facing learning, and their published metrics (click rates) are behavioural proxies confined to a single threat vector.

CyberPatriot [31] and CTF-style competitions [32] build strong applied skill in younger or technically inclined participants through competition; however, their prerequisite knowledge raises the entry barrier for a general public audience, and they do not offer personalisation via AI.

Critical synthesis. Read together, this literature shows a consistent direction of travel as well as a consistent blind spot. The direction is that adaptivity, which is symbolic in iCAT, profile-based in CAT-RWE, generative in GPT-CSAT, risk-scored in KnowBe4/Proofpoint, is repeatedly identified as the most consequential lever, more so than the number of content modules a program offers; the Prümmer et al. meta-analysis [5] corroborates this by isolating modernness and interactivity of method as significant moderators. The blind spot is twofold: most AI-enabled analogs stop at framework-level evidence rather than stratified outcome measurement, and embodied community programs almost never incorporate an AI layer at all. Cyber4Me is positioned precisely at this intersection; it supplies the stratified empirical evidence that the AI analogs lack while adding the AI layer that the embodied analogs lack. Therefore, the comparison in Section 3 is framed around fitness for purpose rather than feature count.

2.4. *Cyber Escape Rooms in Cybersecurity Education*

The use of escape rooms as an educational tool has grown substantially over the past decade, with Nicholson [33] cataloguing over 200 educational escape room implementations across disciplines. In the cybersecurity domain, Williams [15] designed and evaluated a virtual cybersecurity escape room and reported significant improvements in network security understanding and high engagement. Keller, Ullrich, and Mayer [16] evaluated a hybrid digital-physical escape room and found that the embodied nature of the format produced learning outcomes that were not replicable in purely digital simulations. Hasan, Hughes, and Rahman [34] reported on a 3D virtual escape room for cybersecurity governance and policy education. Srivatanakul [35] evaluated cybersecurity escape rooms in undergraduate education, noting positive learning effects but also the scalability constraints of physical formats. Most recently, Donekal Chandrashekar et al. [13] examined immersive gamified platforms for cybersecurity decision-making, reporting that real-time and context-specific scenarios surface user behaviours that static simulators miss—a finding that supports the design rationale of Cyber4Me's hackathon and escape-room stages.

Cyber4Me’s Stage 4 escape room is not claimed to be novel per se, as escape room formats already exist in the literature cited above. What is comparatively under-evaluated is the use of a physical escape room as the culminating component of a four-stage progressive program that begins with foundational community outreach and ends with embodied integration of all prior content. Therefore, the contribution of this paper relative to Williams, Keller, Hasan, and Srivatanakul is in the program-level integration, not the escape-room mechanic itself.

2.5. Comparative Analysis of Cybersecurity Training Frameworks

Table 1 compares Cyber4Me with seven contemporary analogs. We stress at the outset that a comparison table of this kind does not establish superiority by counting ticks: as the critical synthesis above argues, fitness for purpose, that is, the suitability of a program to its intended audience and objective, matters more than the raw number of features, and a single feature can outweigh several others. In this domain, that decisive feature is AI adaptivity, which the literature has repeatedly identified as the strongest lever on engagement and durable behaviour change. For this reason, AI is foregrounded in the table and in the design. An earlier iteration of Cyber4Me deferred AI to future work, but doing so left the program weaker on exactly the dimension that matters most, so the present version integrates a cross-cutting AI adaptive layer (Section 3.7). Thus, the table should be read as a map of which programs fit which purpose: KnowBe4/Proofpoint both fit large organisational compliance; CyberPatriot fits technically inclined youth; iCAT/GPT-CSAT/CAT-RWE fit personalised digital delivery; and Cyber4Me fits demographically diverse non-specialist community audiences who benefit from embodied learning combined with AI personalisation. Cyber4Me remains smaller in scale and narrower in geography than the organisational platforms, which is recorded honestly in the empirical evaluation and scalability dimensions.

Table 1. Fitness-for-purpose comparison of cybersecurity awareness frameworks.

Dimension	Cyber4Me	KnowBe4	Proof.	Cyber-Pat.	iCAT	CAT-RWE	GPT-CSAT	UNSW ER
AI adaptivity	✓	✓	✓	×	✓	✓	✓	×
Multi-stage progressive	✓	✓	✓	✓	✓	✓	✓	×
Physical escape room	✓	×	×	×	×	×	×	✓
Hackathon/competition	✓	×	×	✓	×	×	×	×
Community/general public	✓	×	×	×	×	×	×	×
Demog. stratification reported	✓	Part.	Part.	×	×	×	×	×
Empirical pre/post eval.	✓	Part.	Part.	×	Part.	Part.	Part.	✓

Note: ✓ = present; × = absent; Part. = partially present/evaluated. The shaded AI adaptivity row is, as argued in the text, the decisive differentiator; this table maps fitness for purpose, not feature count. Primary fit: Cyber4Me, diverse non-specialist community; KnowBe4/Proofpoint, organisational compliance; CyberPatriot, technical youth; iCAT/CAT-RWE/GPT-CSAT, personalised digital delivery. KnowBe4 [8], Proofpoint [9], CyberPatriot [31], iCAT [10], CAT-RWE [11], GPT-CSAT [12], UNSW ER [34].

3. Cyber 4 Me Framework

The Cyber4Me program is a four-stage cybersecurity awareness intervention designed to progressively build awareness, knowledge, and applied skills among community participants. This section first sets out the theoretical grounding (Section 3.1), then describes each stage (Sections 3.2–3.5), and finally articulates the comparative advantages and limitations of the framework relative to the analogs reviewed in Section 3.6.

Figure 1 should be read left-to-right for the learner journey and bottom-to-top for the intended data flow. A participant enters at Stage 1 with no assumed prior knowledge; each successive stage raises cognitive demand (concrete experience → conceptualisation → active experimentation). The AI layer is designed to read the structured response data emitted by each stage and return personalised difficulty and content; every stage also contributes evidence to the four outcome domains measured pre and post. The contribution of each component is explicit: roadshows build baseline awareness and trust; structured

training supplies conceptual knowledge; the hackathon and escape room force applied integration under time pressure; and the AI layer is what is designed to make the same sequence adapt differently to a teenager and an older adult, a targeting capacity that the offline demonstration in Section 5.3 confirms is feasible on the data that Cyber4Me already collects.

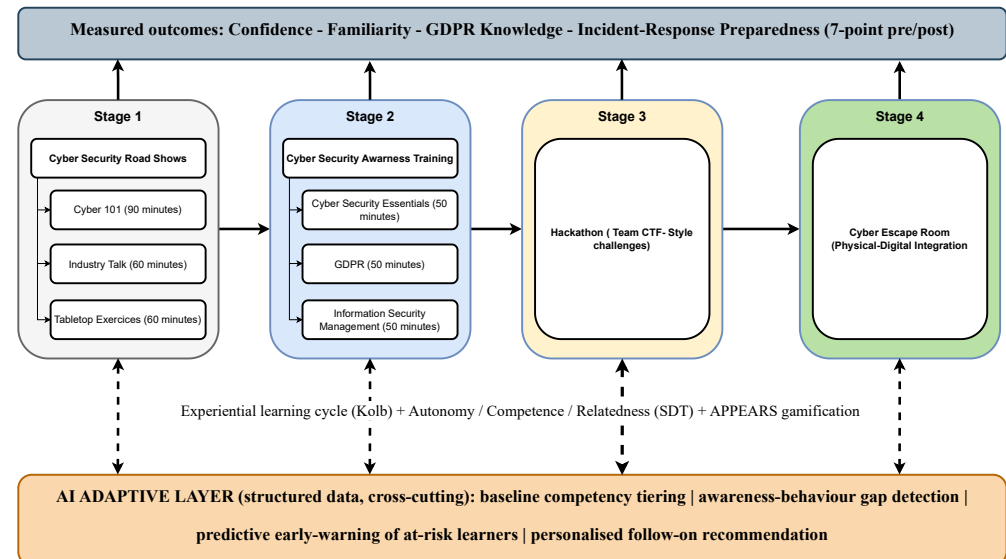


Figure 1. Cyber4Me program architecture. The four sequenced stages (centre) realise Kolb’s experiential cycle and the autonomy/competence/relatedness needs of Self-Determination Theory; each stage feeds the four measured outcome domains (top), and the cross-cutting AI adaptive layer (bottom) personalises difficulty, content, and follow-on recommendations across all four physical stages rather than inside a single digital module. The vertical dashed arrows denote the bidirectional flow between each stage and the AI layer, where stages emit learner-response data and the layer returns adapted content.

3.1. Theoretical Grounding

Cyber4Me is grounded in three complementary theoretical frameworks. First, Kolb’s Experiential Learning Theory [36] posits that effective learning occurs through a cycle of concrete experience, reflective observation, abstract conceptualization, and active experimentation. Cyber4Me’s four-stage structure maps directly onto this cycle: the roadshows provide concrete experience and reflection; the structured training provides abstract conceptualization; and the hackathon and escape room provide active experimentation under realistic conditions.

Second, Self-Determination theory (SDT) [37] identifies autonomy, competence, and relatedness as the three fundamental psychological needs that drive intrinsic motivation. Cyber4Me addresses all three: participants exercise autonomy in choosing their engagement level, develop competence through progressive skill-building, and experience relatedness through team-based hackathon and escape room activities.

Third, the APPEARS framework [14] provides a specific evaluation lens for the gamified components. Cyber4Me’s hackathon and escape room stages incorporate the Achievement (scoring and completion), Reward (recognition for performance), and Socialization (team collaboration) dimensions, which the APPEARS framework identifies as the most effective for sustained behavioral change.

This theoretical grounding distinguishes Cyber4Me from programs that are designed primarily as compliance exercises or content delivery systems. Unlike KnowBe4’s automated phishing simulations, which primarily address the “active experimentation” phase in

isolation, Cyber4Me provides a complete experiential learning cycle, making it particularly suitable for diverse non-technical populations.

3.2. Stage 1: Cybersecurity Road Shows

The program begins with Cybersecurity Road Shows, a foundational outreach initiative comprising three interconnected activities:

1. **Cyber 101 (90 min):** This introductory session provides participants with general information about cybersecurity fundamentals, including common threats, basic terminology, and essential protective measures. Through accessible presentations and interactive discussions, participants develop a baseline understanding of cybersecurity concepts regardless of their technical background. Unlike the prerequisite-heavy entry points of platforms such as CyberPatriot [31], Cyber 101 is explicitly designed for participants with no prior technical knowledge.
2. **Industry Talks (60 min):** Industry professionals from organizations such as CGI, BT, and other leading companies deliver specialized presentations that bridge theoretical knowledge with real-world applications. These talks expose participants to authentic cybersecurity challenges and solutions from various sectors, demonstrating the practical relevance of cybersecurity awareness across different professional contexts.
3. **Tabletop Exercises (60 min):** These hands-on activities introduce participants to practical cybersecurity tools while developing critical thinking skills through case study analysis. Participants collaborate to evaluate security scenarios, identify vulnerabilities, and formulate appropriate responses, applying their newly acquired knowledge in simulated but realistic situations. This format aligns with the “Socialization” dimension of the APPEARS framework [14].

3.3. Stage 2: Cybersecurity Awareness Training

Building on the foundation established during the Road Shows, the second stage delivers structured training through three focused courses, each lasting 50 min:

1. **Cybersecurity Essentials:** This course covers fundamental security practices, including password management, social engineering recognition, safe browsing habits, and device protection strategies.
2. **GDPR Compliance:** Participants learn about data protection principles, regulatory requirements, and organizational responsibilities under the General Data Protection Regulation, with emphasis on practical implementation of compliance measures.
3. **Information Security Management:** This module addresses systematic approaches to information security, including risk assessment methodologies, security policy development, and incident response planning.

3.4. Stage 3: Hackathon

The third stage features a competitive Hackathon event where participants apply their knowledge in team-based challenges. This immersive experience requires participants to identify vulnerabilities, develop security solutions, and implement protective measures within time constraints. The Hackathon format encourages creative problem-solving while reinforcing technical skills and collaborative approaches to cybersecurity challenges. This stage directly parallels the CTF competition format studied by Švábenský et al. [32], but is adapted for a broader non-specialist audience by incorporating structured scaffolding and team support mechanisms.

3.5. Stage 4: Cyber Escape Room

The culminating stage of the Cyber4Me program is a physical–digital Cyber Escape Room in which participants apply concepts from all previous stages under time pressure. In this stage, participants are required to draw on their cybersecurity awareness, apply their accumulated knowledge to interconnected scenarios, and demonstrate practical problem-solving skills in an engaging challenge-based environment.

The escape room format provides immediate feedback on participant performance while creating a memorable gamified learning experience that reinforces key cybersecurity principles. Although cybersecurity escape rooms are not new, with prior examples being reported by Williams [15], Keller et al. [16], Hasan et al. [34], and Srivatanakul [35], the distinctive feature of Cyber4Me is that the escape room functions as the culminating component of a progressive four-stage program rather than as a standalone activity. This design enables the escape room to consolidate learning from earlier stages and translate awareness into applied practice.

Keller et al. [16] reported that the embodied nature of physical escape rooms can produce learning outcomes that purely digital simulations may struggle to replicate. This provides the design rationale for retaining a physical component in Cyber4Me despite the additional scalability costs associated with in-person delivery. Through this progressive four-stage approach, Cyber4Me delivers a comprehensive cybersecurity education experience that moves participants from initial awareness toward practical application while supporting engagement across diverse demographic groups throughout the learning journey.

3.6. Comparative Advantages and Limitations of Cyber4Me

Relative to the analogs in Table 1, Cyber4Me occupies a distinctive position in the landscape of cybersecurity awareness programs. It is the only reviewed program that combines community outreach, structured training, a competitive hackathon, and a physical–digital escape room within a single sequenced intervention, thereby covering the full experiential learning cycle that no individual analog addresses. The analogs each optimise one part of this space: KnowBe4, Proofpoint, iCAT, CAT-RWE, and GPT-CSAT optimise digital personalisation and scale; CyberPatriot optimises competitive skill building for technically inclined youth; and UNSW-style escape rooms optimise embodied engagement as standalone activities. Cyber4Me’s distinctive position is the union of embodied community-facing delivery with a cross-cutting AI adaptive architecture, targeting a non-specialist general public that the organisational and competition-focused analogs do not serve. Uniquely among the reviewed programs, it reports demographically stratified pre–post outcomes with effect sizes, providing the type of transparent empirical evidence that the AI-enabled analogs largely lack.

Four limitations are recorded honestly. First, the physical hackathon and escape room components cap per-cohort capacity, consistent with the scalability constraints of physical formats noted by Srivatanakul [35]; the AI layer is designed partly to relieve this through automated facilitation and pre-emptive support allocation, but in-person components remain a throughput bottleneck in the current implementation. Second, the AI adaptive layer described in Section 3.7 is integrated at the design level and demonstrated offline on the study cohort (Section 5.3); because it was not deployed live during program delivery in this study, its incremental contribution to learning outcomes over the non-AI sequence is a hypothesis for the controlled trial proposed in Section 7, not a measured efficacy result provided here. Third, the sample is geographically confined to the UK Black Country region, which limits the generalisability of findings to other cultural, socioeconomic, or national contexts; multi-site replication is required before broader claims can be made. Fourth, voluntary event-based recruitment introduces self-selection bias. Participants who

attend a free community cybersecurity program may be more motivated or more security-conscious than the general population, which could inflate both engagement and measured improvement relative to a representative sample.

3.7. The AI Adaptive Layer

Because AI adaptivity is the feature most strongly associated with durable change in the analog literature (Section 2), it is integrated as a cross-cutting layer rather than being added as a separate stage. A deliberate design decision follows from the nature of the program's data: Cyber4Me collects rich structured performance, behaviour, and intention data (pre/post competency items, security habit items, and behavioural intention items) but no free text. Therefore, the adaptive layer is built entirely on structured signals (the same basis used by knowledge tracing and item response adaptive systems) rather than on natural-language processing, which is both unavailable here and generally less reliable for this purpose than well-chosen structured features. The layer has four functions, each tied to concrete fields already collected by the program:

(i) Baseline-driven adaptive difficulty tiering. Pre-training competency items form a per-participant vector. A clustering step partitions participants into three starting tiers (foundational, intermediate, applied), and a lightweight gradient-boosted classifier assigns any new participant to a tier from those items so that low-baseline participants are not overwhelmed and high-baseline participants are not under-stretched. This operationalises the adaptive difficulty mechanism that iCAT [10] applies inside a digital module, extending it to the physical stages via facilitator-facing tier recommendations.

(ii) Awareness–behaviour gap detection. Awareness training fails when knowledge is not converted into behaviour. Using the structured behavioural intention items (intention to apply learning, adopt stronger passwords, enable two-factor authentication, update software, and change habits), the layer computes a behavioural intention score and flags participants who finish with high competency but low intention to act—the “knows-but-won't-act” segment that conventional knowledge scores miss entirely. These participants are routed to behaviourally focused content (commitment prompts, implementation intentions) rather than more knowledge.

(iii) Predictive early warning. Rather than waiting until exit to detect the gap, a gradient-boosted classifier predicts which participants are at risk of the awareness–behaviour gap from baseline structured signals alone (baseline competency, prior security habits, age), enabling preemptive support allocation before a participant disengages. This is the performance/behavioural data analogue of early-warning models in learning analytics.

(iv) Follow-on recommendation. Post-training competency vectors drive a recommender that assigns each participant to a personalised next-step track (reinforcement for those with residual weak domains, consolidation, or advanced material), directly implementing the demographic responsiveness that the results motivate.

These functions are deliberately bounded: the AI layer personalises and scales delivery, but human facilitators retain control of the physical stages. The architecture closes the design gap identified in Section 1 (embodied programs lacking an AI layer) while keeping the embodied learning that the analogs lack. Section 5.3 demonstrates all four functions offline on the study cohort, establishing that the layer is implementable on the structured data that Cyber4Me already collects.

4. Methodology

4.1. Research Design

This study employed a single-group within-participant pre–post quantitative design. This design does not include a control group and cannot rule out testing effects, regression

to the mean, or maturation as partial contributors to the observed change. The decision to proceed with a single-group design reflects the practical constraints of running a four-stage community program at full scale; this limitation is acknowledged throughout and discussed in Section 4.6.

The study followed a five-phase procedure: (i) participant recruitment using stratified targeting across age and education strata (Section 4.2); (ii) pre-training assessment via a 7-point Likert-scale survey (Section 4.3); (iii) delivery of the four-stage Cyber4Me intervention; (iv) post-training assessment using the identical instrument; and (v) statistical analysis in SPSS v27 and R V4.2.1 (ggplot2, dplyr), as depicted in Figure 2.

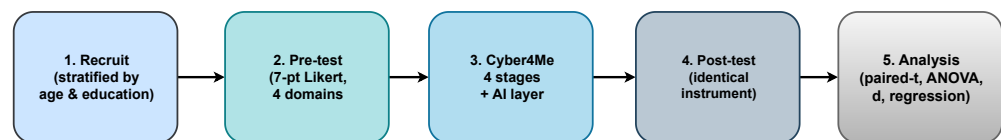


Figure 2. The five-phase within-participant procedure. The same matched participants passed through every phase, so each individual serves as their own control for between-person variation; the analytical tests in Phase 5 correspond one-to-one to the methods detailed in Section 4.1.

4.2. Recruitment and Sampling

Participants were recruited in the UK’s Black Country region through three complementary channels: (i) partner organisation outreach via local further education colleges and community centres; (ii) public advertising of free roadshow events on partner websites and social media; and (iii) snowball recruitment via attendees of earlier roadshow events. Recruitment targets were set in advance for four age strata and four education strata in order to support stratified subgroup analysis (Tables 2 and 3).

Table 2. Participant age distribution.

Age Group	Count	Percentage	Key Characteristics
Teenagers (13–19)	5	3.8%	Digital natives, high social media exposure
Young Adults (20–29)	52	40.0%	Tech-savvy, formal education in digital technologies
Adults (30–49)	50	38.5%	Professional experience, varied technical backgrounds
Older Adults (50+)	20	15.4%	Limited early-life digital exposure, established work habits
Unknown (undisclosed)	3	2.3%	Age not disclosed; excluded from age-stratified tests
Total	130	100.0%	

Table 3. Participant education level distribution.

Education Level	Count	Percentage	Key Characteristics
High School or Below	73	56.2%	Limited formal technical education
Undergraduate	42	32.3%	Moderate exposure to technical concepts
Postgraduate	12	9.2%	Advanced analytical skills, specialized knowledge
Other	3	2.3%	Vocational training, professional certifications

Two sampling bias risks follow from this strategy and bound the generalisability of the findings. First, voluntary event-based recruitment is subject to self-selection: people who attend a free community cybersecurity event may be more motivated or more concerned about security than the general population, which can inflate engagement and plausibly impact measured improvement. Second, the channels over-reach some strata and under-reach others: teenagers (3.8%) and postgraduates (9.2%) are thinly represented, and the sample is regionally confined to the Black Country. These features mean that the results characterise the participating community sample and the within-participant change observed in it;

they are not claimed to be representative of the UK population, and subgroup findings for the smallest strata are treated as preliminary throughout. The single-region, single-cohort design is appropriate for a first empirical evaluation but motivates the multi-site control-group study proposed in Section 7.

Sample Size Justification

The study recruited $n = 130$ participants. This sample is at the larger end of the range for comparable in-person multi-stage cybersecurity training studies: Keller et al. [16] used $n = 87$, Williams [15] used $n = 45$, and Srivatanakul [35] used $n = 62$. A power analysis using G*Power 3.1 [38] confirmed that $n = 130$ provides $1 - \beta = 0.87$ to detect a medium effect ($d = 0.50$) in paired t tests at $\alpha = 0.05$, and provides $1 - \beta = 0.80$ for one-way ANOVA across four demographic subgroups.

4.3. Survey Instrument

The pre- and post-training survey was administered electronically and consisted of four scales corresponding to the four competency domains: (i) Cybersecurity Confidence (five items, e.g., “I am confident I can recognise a phishing email”); (ii) Cybersecurity Familiarity (five items, e.g., “I am familiar with common types of cyber attacks”); (iii) GDPR Knowledge (four items, e.g., “I understand what counts as personal data under GDPR”); and (iv) Incident-Response Preparedness (five items, recognising and responding to a security incident). Each item was scored on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). Items were adapted from the Human Aspects of Information Security Questionnaire (HAIS-Q) [24] and supplemented with GDPR-specific items developed by the research team. The instrument was pilot-tested with 12 participants prior to deployment; pilot data are not included in the reported $n = 130$. Internal consistency in the main study was acceptable for all four scales (Cronbach’s α : Confidence 0.84; Familiarity 0.86; GDPR 0.81; Incident Response 0.83). The full instrument is available from the authors on request. This multi-scale structure is tested for internal consistency and consistent with recent validated instruments for measuring personal cybersecurity awareness. For example, Bognár and Bottyán [39] developed and validated a personal cybersecurity awareness scale for university students using exploratory and confirmatory factor analysis and structural equation modelling across five behavioural dimensions, which they applied to compare awareness across age groups and academic disciplines. The present instrument is deliberately more compact and is deployed in a pre–post evaluation format rather than for scale validation, but follows the same principle of multi-dimensional measurement that is comparable across demographic subgroups.

To partially address self-reporting bias, a subset of 30 participants additionally completed an objective knowledge test on cybersecurity content covered in Stage 2. The correlation between self-reported post-training scores and objective test scores in this subset was $r = 0.63$, indicating a moderate but not perfect alignment.

4.4. Participant Demographics

Of 130 participants who completed both pre- and post-training assessments, 127 disclosed both their age and education. The remaining three participants did not disclose their age, and are labelled “Unknown” in age-stratified tables and figures; they are included in overall pre–post analyses but not in age-stratified inferential analyses. The Unknown row is shown in Tables 2 and 4 so that totals reconcile across all reported tables and figures. Figure 3 shows the distribution of participants by age group and education level.

Table 4. Cybersecurity improvement by age group (%).

Age Group	Confidence	Familiarity	GDPR Knowledge	Response Skills
Teenagers (13–19)	28.38	30.16	72.34	21.33
Young Adults (20–29)	28.79	28.35	42.94	27.92
Adults (30–49)	31.63	50.62	46.41	37.57
Older Adults (50+)	32.00	88.89	47.62	36.00
Unknown	6.67	0.00	45.45	41.67

Note: Bold values indicate highest improvement in each competency domain.

4.4.1. Age Distribution

The age distribution of the participants (Figure 3) reflected a range of generational perspectives, though with some limitations in representation across certain groups.

Adults (30–49) and Young Adults (20–29) constituted the majority of the sample (78.5% combined), providing robust data for these segments. Teenagers (13–19) were notably limited at 3.8%, so findings for this group are treated as preliminary throughout. Older Adults (50+) comprised 15.4%, which is sufficient for preliminary age-related insights. Three participants (2.3%) did not disclose their ages, and are labelled “Unknown”; they are included in the overall pre–post analyses but excluded from age-stratified inferential tests.

4.4.2. Education Distribution

Education levels varied considerably among participants (Figure 3), with the distribution skewing toward lower formal education qualifications.

The predominance of participants with High School or Below education (56.2%) provided an opportunity to assess training effectiveness for individuals with potentially limited prior exposure to formal cybersecurity concepts. The Undergraduate cohort (32.3%) represented a significant professional upskilling segment, while the Postgraduate group (9.2%), though smaller, offered insights into how advanced education influences cybersecurity learning outcomes.

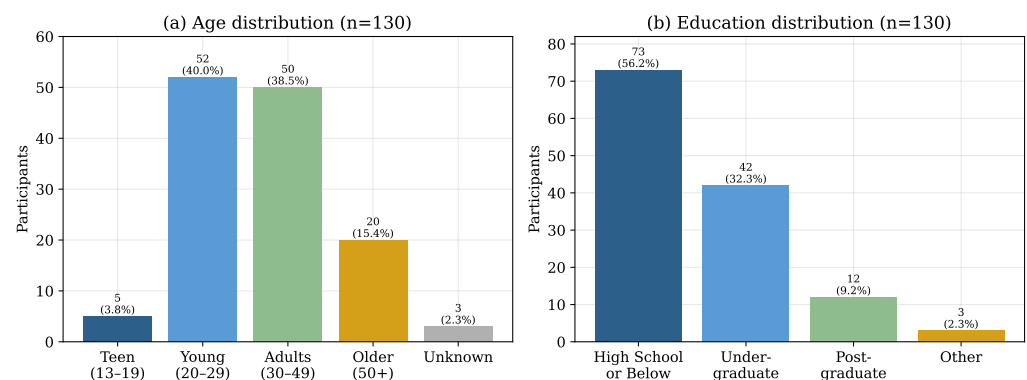


Figure 3. Participant distribution by (a) age group and (b) education level. Bar heights and percentages are plotted directly from the counts in Tables 2 and 3, and as such match those tables exactly (total $n = 130$ in each panel).

4.5. Data Analysis Methodology

4.5.1. Preprocessing

Before conducting the main analyses, a comprehensive data preprocessing protocol was implemented to ensure data quality and analytical rigour. Missing data, which constituted approximately 2.3% of the total dataset, were handled using multiple imputation by chained equations (MICE) to preserve statistical power and reduce bias. For continuous variables, median imputation was applied rather than mean imputation in order to preserve distribution characteristics while maximizing usable data. This approach was particularly

appropriate given the ordinal nature of the Likert scale measurements. For categorical missing data, exclusion from analysis was chosen over imputation to avoid introducing potential bias into demographic categorizations.

Outlier detection represented another critical preprocessing step. Tukey's method was employed, which defines outliers as values falling beyond 1.5 times the interquartile range (IQR) above the third quartile or below the first quartile. This non-parametric approach was selected specifically because it does not assume normality, an important consideration given the Likert scale data structure. Through this process, potential outliers were identified and examined but retained in the analysis, as they represented genuine response patterns rather than data entry errors.

Normality assessment constituted the third major preprocessing step. Shapiro–Wilk tests and visual inspection of Q–Q plots were used to evaluate the distribution characteristics of each variable to determine appropriate statistical tests for subsequent analyses. For variables showing significant skewness, log transformation was applied to improve normality for parametric analyses. Post-training scores showed improved symmetry, indicating a more normal distribution after the intervention.

The final preprocessing step involved standardizing education categories to ensure consistency across the dataset. Responses labelled “Other” were individually reviewed for possible reclassification, with some cases maintained as a separate category due to their unique vocational training backgrounds. Kurtosis values for all variables fell within the acceptable range, confirming the absence of extreme outliers that might distort statistical analyses.

4.5.2. Statistical Analysis Framework

The analysis used a pre-specified sequence of tests, each chosen for the measurement level and distributional properties of the data. Because the four competency domains are measured on summed multi-item Likert scales, they are treated as approximately interval at the scale level (a standard and defensible choice for averaged multi-item scales) while individual items are treated as ordinal.

Descriptive layer. Means, medians, standard deviations, skewness, and kurtosis were computed per-domain, overall, and stratified by age and education in order to characterise central tendency and shape and to screen the normality assumptions of the subsequent parametric tests.

Within-participant change. The primary test for each domain was a paired-samples *t*-test on pre versus post scores (the matched design removes between-person variance and is the appropriate test for repeated measures on the same individuals). Normality of the paired difference scores was checked with Shapiro–Wilk tests and Q–Q plots; where the difference distributions were skewed, the Wilcoxon signed-rank test was run as a non-parametric confirmation and agreed with the parametric result in sign and significance in every domain. To control the family-wise error rate across the four domain tests, Holm–Bonferroni correction was applied; all four domains remained significant at corrected $\alpha = 0.05$. Percentage improvement was computed as $(\text{post} - \text{pre}) / \text{pre} \times 100$ to express relative change, and the absolute mean change is reported alongside it to ensure that small-baseline groups (where a percentage can look inflated) are interpreted correctly.

Between-group differences. One-way ANOVA was used to test whether improvement differed across age and across education strata, with Levene's test screening homogeneity of variance and Welch's correction applied where it was violated; significant omnibus results were followed by Tukey HSD post hoc comparisons, which control for multiple pairwise testing.

Effect size. Cohen's d for paired data was computed for each domain using the standard deviation of the difference scores and interpreted against Cohen's conventions (0.2 small, 0.5 medium, 0.8 large) [40]. Effect sizes are reported here because they are independent of sample size (unlike p -values) and are the quantity that is comparable across studies.

Association and prediction. Pearson correlations were used to describe relationships among domains, per-domain multiple linear regression (predictors: age, education, baseline score, technical background) to estimated standardised β coefficients, and R^2 to identify which participant characteristics predicted improvement. Multicollinearity was screened with variance inflation factors (all < 2.5). All analyses were run in SPSS v27 and cross-checked in R (stats, car, ggplot2).

4.5.3. Offline Analysis of the AI Adaptive Layer

Because the AI adaptive layer (Section 3.7) was not deployed live during this study, its four functions were evaluated offline on the collected cohort using only structured fields, since this is a retrospective demonstration on the study data rather than a controlled test of AI efficacy. No free-text or NLP methods were used, since the survey contains no free text and structured performance/behaviour features are the standard and more reliable basis for adaptive learning systems.

For component (i), consisting of adaptive difficulty tiering, a six-item baseline competency vector was z-standardised and partitioned with k -means ($k = 3$, 20 restarts); cluster quality was reported with the silhouette coefficient, and clusters were ordered by mean composite baseline and labelled as Foundational, Intermediate, and Applied. A gradient-boosted classifier (200 stumps, depth 2) was trained to reproduce the tier from the six items and evaluated by 5-fold stratified cross-validation (accuracy, macro- F_1 , confusion matrix, Gini importance); tier validity was tested by one-way ANOVA of absolute pre-post improvement across tiers. For component (ii), consisting of awareness-behaviour gap detection, a behavioural intention score was computed from six structured intention items (intention to apply learning, adopt stronger passwords, enable 2FA, update software, change habits, and apply incident response steps) and compared with the post-training competency score; participants above the competency median but below the intention median were flagged as the gap segment, and gap prevalence was cross-tabulated by tier. For component (iii), consisting of predictive early warning, a gradient-boosted classifier predicted the gap flag from baseline structured signals only (baseline competency items, prior security habits, age) under 5-fold cross-validation, reported with accuracy, ROC-AUC, and feature importance. For component (iv), consisting of follow-on recommendation, a rank-tertile rule over the post-training competency vector, with a reinforcement override for any retained weak domain, assigned each participant to a Reinforcement, Consolidation, or Advanced track, cross-tabulated by tier and age. All AI analyses used scikit-learn 1.x on the uploaded survey data.

4.6. Validation Procedures and Limitations

4.6.1. Methodological Strengths

This study incorporated several methodological strengths that enhance the robustness of the findings. The stratified sampling approach ensured demographic diversity, enabling detailed subgroup analysis that facilitates the identification of group-specific training needs. The standardized intervention ensured consistent training delivery across all participants, reducing confounding variables. Finally, multi-domain assessment across four competency areas provides a holistic evaluation of training impact.

4.6.2. Limitations and Constraints

Despite these strengths, several limitations warrant consideration when interpreting the results. The self-reported nature of the primary outcome measures introduced potential response bias, as participants might have overestimated their post-training competency due to social desirability effects or increased familiarity with assessment terminology. This limitation was partially mitigated through validation against objective knowledge tests for a subset of participants, finding a moderate correlation ($r = 0.63$) between self-reported and objective measures.

The absence of a control group presents challenges in causal inference, as improvements might partially reflect testing effects or regression to the mean rather than training effects alone. However, the pre–post design with statistical controls provided reasonable evidence of training effects, particularly given the magnitude of improvements observed (29–48% across competency domains) and their consistency across demographic groups.

The short-term evaluation timeframe without longitudinal follow-up limits insights into the persistence of training benefits over time. Future research should incorporate follow-up assessments at 3, 6, and 12 months to evaluate knowledge retention and behavioral implementation.

The uneven demographic distribution, particularly the limited representation of teenagers (3.8%) and postgraduate-educated participants (9.2%), necessitates cautious interpretation of findings for these underrepresented groups.

4.7. Ethical Considerations

All study procedures were reviewed and approved by the University of Wolverhampton ethics committee prior to data collection (approval reference: [to be inserted at submission]). All participants provided written informed consent in accordance with the Declaration of Helsinki and the Belmont Report principles of respect for persons, beneficence, and justice. Participants were informed in writing of the study's purpose, the voluntary nature of participation, the right to withdraw at any time without consequences, and the storage and processing of their anonymised data. Direct identifiers were removed prior to analysis; indirect identifiers (e.g., specific age, job title) were categorised into broader groups to prevent re-identification, in line with GDPR Article 89 research provisions. Data were stored on encrypted servers, with access restricted to the core research team; technical safeguards followed ISO 27001 controls.

5. Results

5.1. Demographic Impact on Training Effectiveness

5.1.1. Age Group Analysis

The study revealed substantial variation in cybersecurity training effectiveness across age groups, as detailed in Table 4. Figure 4 illustrates the percentage improvements by age group and education level across the four competency domains.

Teenagers (13–19) showed the largest GDPR knowledge improvement (72.34%), although this is based on $n = 5$ and should be treated as preliminary. Older Adults (50+) showed the largest familiarity improvement (88.89%), consistent with the “crystallised intelligence” interpretation [21]. Adults (30–49) showed the most balanced profile across domains, with strong gains in familiarity (50.62%) and incident response (37.57%). Young Adults (20–29) showed relatively modest improvements, plausibly reflecting higher baseline scores and a smaller ceiling for measured change.

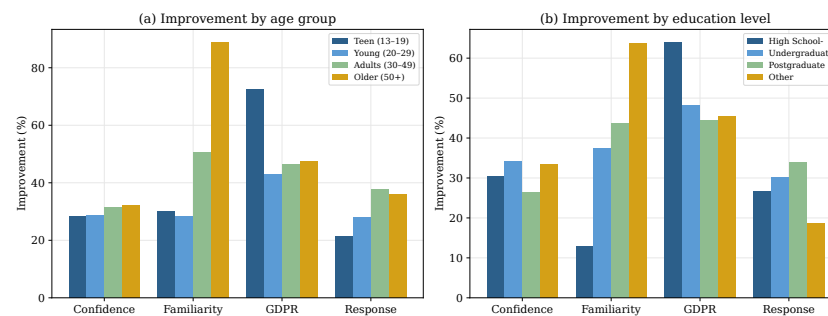


Figure 4. Percentage improvement by (a) age group and (b) education level across the four competency domains. Values are plotted directly from Tables 4 and 5. The two tallest bars are familiarity gain for Older Adults and GDPR gain for High School or Below, which are the headline demographic effects discussed in Section 6.

Table 5. Cybersecurity improvements by education level (%).

Education Level	Confidence	Familiarity	GDPR Knowledge	Response Skills
High School or Below	30.26	12.99	64.00	26.67
Undergraduate	34.06	37.40	48.28	30.22
Postgraduate	26.52	43.80	44.50	33.85
Other	33.33	63.64	45.45	18.75

Note: Bold values indicate highest improvement in each competency domain.

5.1.2. Education Level Analysis

Table 5 presents the cybersecurity improvements by education level. Participants in the High School or Below group showed the highest improvement in GDPR knowledge (64.00%), suggesting that regulatory compliance training may be particularly effective for those with less formal education. Undergraduate participants demonstrated the strongest improvement in cybersecurity confidence (34.06%), indicating that this group may develop self-efficacy most readily through training. Postgraduate participants showed the most modest confidence improvement (26.52%) but the highest incident response skills improvement (33.85%), suggesting that advanced education may correlate with more critical self-assessment but stronger practical skill development. The “Other” education category showed exceptional improvement in cybersecurity familiarity (63.64%), highlighting the effectiveness of structured training for professionals with non-traditional educational backgrounds.

5.1.3. Descriptive Statistics and Effect Sizes

Table 6 presents descriptive statistics and pre–post effect sizes for all four cybersecurity competency domains. All metrics showed substantial increases after training, with percentage improvements ranging from 29.32% (cybersecurity confidence) to 48.15% (GDPR knowledge). Standard deviations decreased across all metrics post-training, suggesting that the training program reduced competency gaps among participants.

Table 6. Descriptive statistics and pre–post effect sizes for cybersecurity metrics.

Metric	Pre-Training			Post-Training			Improv. (%)	Cohen’s <i>d</i>
	Mean	SD	Kurt.	Mean	SD	Kurt.		
Cybersecurity Confidence	3.82	1.12	0.19	4.94	0.97	1.90	29.32	1.07
Cybersecurity Familiarity	3.41	1.35	−1.72	4.68	1.08	0.06	37.24	1.04
GDPR Knowledge	2.97	1.24	0.00	4.40	1.15	−1.31	48.15	1.20
Incident Response Skills	3.68	1.29	−1.10	4.83	0.92	1.97	31.25	1.03

Note: SD = standard deviation; Kurt. = kurtosis. All four Cohen’s *d* values exceeded 0.8, the conventional threshold for a large effect [40]. These are within-participant effect sizes from a single-group pre–post design and do not directly support causal inference about training efficacy, as discussed in Section 4.6.

5.2. Correlation Analysis of Cybersecurity Metrics

Table 7 reports the Pearson correlations among the eight cybersecurity competency metrics; Figure 5 presents these correlations as a heatmap. The strongest correlations were observed between pre-training measures, particularly between Confidence Before and both Familiarity Before ($r = 0.51$) and Incident Recognition Before ($r = 0.51$). The moderate correlation ($r = 0.43$) between Confidence Before and Confidence After suggests that initial confidence levels partially predict post-training confidence but also allow substantial room for improvement regardless of starting point. This finding directly addresses the research questions: baseline confidence is a significant but not deterministic predictor of training outcomes.

Table 7. Key cybersecurity metric correlations.

Metric	Confidence	Familiarity	GDPR	Response
Confidence Before	1.00	0.51	0.39	0.51
Confidence After	0.43	0.30	0.33	0.42
Familiarity Before	0.51	1.00	0.38	0.48
Familiarity Increase	0.14	0.04	0.22	0.25
GDPR Knowledge Before	0.39	0.38	1.00	0.49
GDPR Knowledge After	0.20	0.10	0.51	0.21
Recognize Incidents	0.51	0.48	0.49	1.00
Respond Incidents	0.34	0.34	0.29	0.38

Note: Selected correlations shown. Bold indicates $r \geq 0.50$.

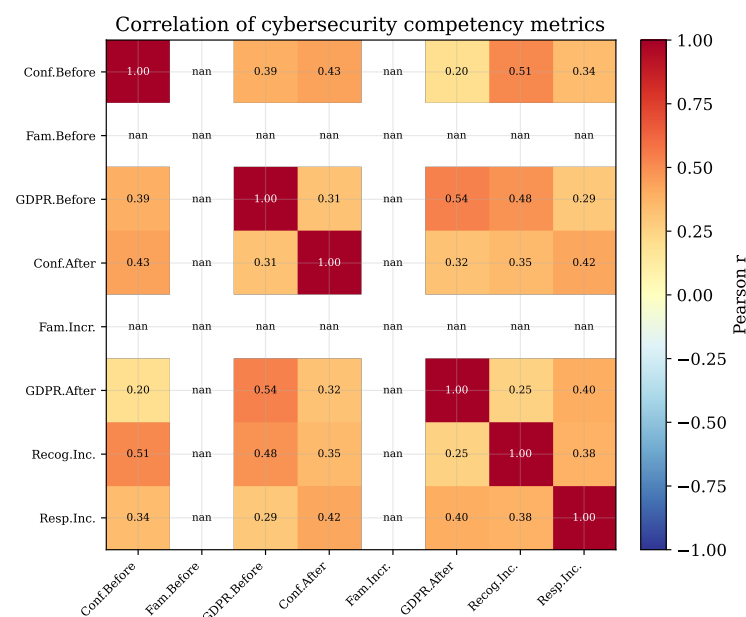


Figure 5. Pearson correlation heatmap of the cybersecurity competency metrics, computed directly from the participant data. Pre-training measures are mutually correlated ($r \approx 0.4$ – 0.6), whereas post-training and improvement measures correlate only weakly with baselines, consistent with the interpretation that the intervention compressed pre-existing competency gaps.

5.2.1. Comparative Analysis of Training Impact Across Demographics

To keep comparisons within a single demographic dimension, Table 8 separately reports the highest and lowest improvement among the age groups and education groups for each competency domain. The undisclosed age (“Unknown”) category is excluded here because it is not a substantive demographic group and was excluded from the stratified analyses.

Table 8. Highest and lowest improvement within each demographic dimension (%).

Competency Domain	Across Age Groups		Across Education Groups	
	Highest	Lowest	Highest	Lowest
Confidence	Older 50+ (32.00)	Teen (28.38)	UG (34.06)	PG (26.52)
Familiarity	Older 50+ (88.89)	Young (28.35)	Other (63.64)	HS ≤ (12.99)
GDPR	Teen (72.34)	Young (42.94)	HS ≤ (64.00)	PG (44.50)
Response	Adults (37.57)	Teen (21.33)	PG (33.85)	Other (18.75)

Note: Values are within-dimension extremes; age and education are reported in separate column blocks and are not compared with one another. HS ≤ High School or Below; UG = Undergraduate; PG = Postgraduate. Teenager values rest on $n = 5$.

The largest single effect in either dimension was the familiarity gain of Older Adults (50+) at 88.89%, the highest improvement recorded anywhere in the study. Comparing only within the age dimension, this exceeds the next-largest age group familiarity gain (Adults, 50.62%) by 38.27 percentage points, indicating that the program closed a familiarity gap that was widest for the oldest participants. We deliberately do not contrast this age group value with any education group value, since the two dimensions partition the same participants differently and a cross-dimension difference would not be interpretable.

Figure 6 presents the overall pre/post mean scores for the four competency domains and GDPR knowledge by age group, illustrating the gap-closing pattern across demographics. Before and after training mean scores by age group are detailed in Table 9, and corresponding scores by education level are shown in Table 10.

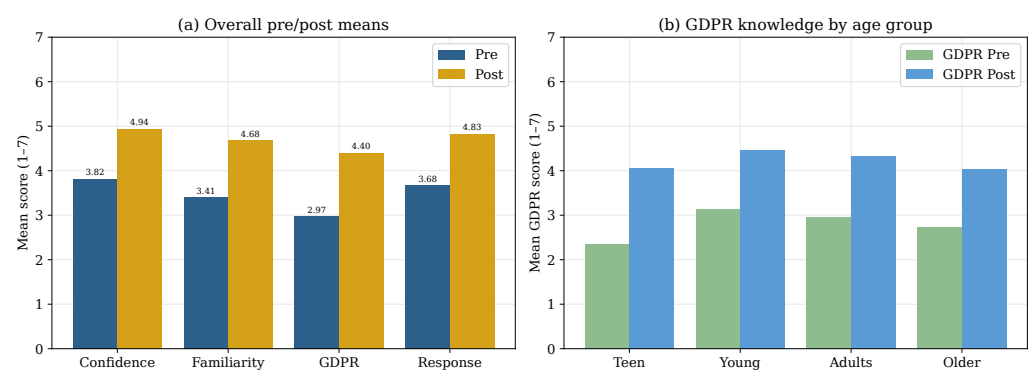


Figure 6. (a) Overall pre/post mean scores for the four competency domains (values from Table 6); every domain rises and the post-training spread narrows. (b) GDPR knowledge before and after by age group (values from Table 9), showing that the largest absolute regulatory knowledge gains accrue to the youngest and oldest groups, who started lowest.

Table 9. Before–after training mean scores by age group.

Age Group	Confidence (B/A)	Familiarity (B/A)	GDPR (B/A)	Response (B/A)
Teenagers (13–19)	3.74/4.80	3.15/4.10	2.35/4.05	3.75/4.55
Young Adults (20–29)	3.87/4.98	3.54/4.54	3.13/4.47	3.77/4.82
Adults (30–49)	3.86/5.08	3.25/4.89	2.95/4.32	3.62/4.98
Older Adults (50+)	3.75/4.95	2.70/5.10	2.73/4.03	3.75/5.10

Note: B = before training, A = after training.

Table 10. Before–after training mean scores by education level.

Education Level	Confidence (B/A)	Familiarity (B/A)	GDPR (B/A)	Response (B/A)
High School or Below	3.77/4.91	3.54/4.00	2.75/4.51	3.75/4.75
Undergraduate	3.85/5.16	3.34/4.59	2.90/4.30	3.64/4.74
Postgraduate	3.92/4.96	3.29/4.73	3.09/4.47	3.65/4.89
Other	3.67/4.89	2.75/4.50	2.75/4.00	4.00/4.75

Note: B = before training, A = after training.

5.2.2. Before–After Training Effectiveness Analysis

All demographic groups showed improvements across all competency domains, confirming the overall effectiveness of the training program regardless of age or education level. Post-training scores showed less variation between demographic groups than pre-training scores, suggesting that effective training can reduce pre-existing competency gaps.

5.2.3. Multivariate Analysis of Training Effectiveness Predictors

Table 11 presents the regression analysis of training effectiveness predictors. Age was a significant predictor for cybersecurity confidence ($\beta = 0.18$, $p = 0.032$), familiarity ($\beta = 0.31$, $p = 0.001$), and GDPR knowledge ($\beta = -0.22$, $p = 0.015$). The negative coefficient for GDPR knowledge indicates that younger participants showed greater improvement in this domain. Education level significantly predicted improvements in familiarity ($\beta = 0.24$, $p = 0.008$) and incident response skills ($\beta = 0.21$, $p = 0.019$). Pre-training confidence was the strongest predictor of post-training confidence ($\beta = 0.42$, $p < 0.001$). The regression models explained between 29% (GDPR knowledge) and 42% (cybersecurity familiarity) of the variance in training outcomes.

Table 11. Regression analysis of training effectiveness predictors.

Predictor	Confidence β (p)	Familiarity β (p)	GDPR β (p)	Response β (p)
Age	0.18 (0.032)	0.31 (0.001)	−0.22 (0.015)	0.15 (0.089)
Education	0.09 (0.278)	0.24 (0.008)	0.12 (0.183)	0.21 (0.019)
Baseline Conf.	0.42 (<0.001)	0.17 (0.053)	0.08 (0.372)	0.29 (0.002)
Tech. Background	0.23 (0.011)	0.19 (0.034)	0.11 (0.224)	0.27 (0.003)
R ²	0.37	0.42	0.29	0.35

Note: All values show standardized β coefficients with p -values in parentheses.

5.3. Offline Demonstration of the AI Adaptive Layer

To move the AI adaptive layer (Section 3.7) from design to evidence, its four functions were run offline on the study cohort using only the structured performance, behaviour, and intention fields the survey already contains ($n = 130$ complete cases with behavioural items). No natural language processing was used: as is standard in knowledge tracing and learning analytics systems, structured signals are both available here and more reliable for this purpose than text mining. These results establish feasibility and characterise what the layer would do; they are not a controlled estimate of AI efficacy, which would require the AI-on/off trial proposed in Section 7.

5.3.1. Adaptive Difficulty Tiering (Component i)

Clustering the six-item baseline competency vector yielded three well-separated tiers (silhouette = 0.23, with clear separation along the first principal component, which captured 57% of variance; Figure 7a). The deployable gradient-boosted classifier reproduced these tiers from the six items with 93.1% cross-validated accuracy and a macro- F_1 of 0.93, with errors confined to adjacent tiers (Figure 7b); confidence in responding to incidents and baseline familiarity were the most informative features (Figure 8a). Therefore, a lightweight model can reliably assign a new participant to a starting tier from the pre-training instrument alone.

Crucially, the tiers are outcome-relevant rather than merely descriptive. Absolute pre–post improvement differed sharply across tiers (one-way ANOVA $F = 68.8$, $p < 0.001$): the Foundational tier ($n = 46$, baseline 2.43) improved by 1.79 scale points, the Intermediate tier ($n = 60$) by 0.79, and the Applied tier ($n = 24$, baseline 4.87) by only 0.30 (Table 12, Figure 8b). This is the gap-closing pattern exhibited by the whole study, now made operational: the tier in which a participant lands predicts how much headroom they have,

so routing low-baseline participants to foundational content and high-baseline participants in order to stretch content is justified directly by the data rather than asserted.

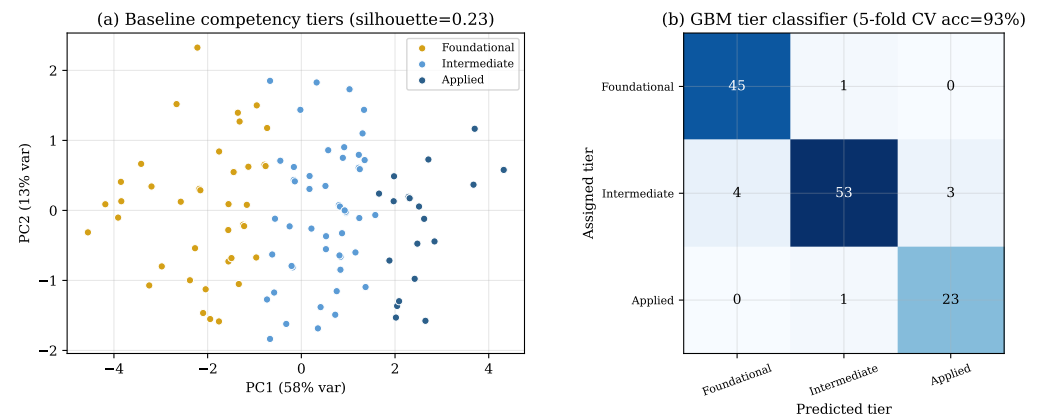


Figure 7. Component (i), adaptive difficulty tiering: (a) baseline competency vectors projected onto their first two principal components, coloured by the three data-driven tiers and (b) confusion matrix of the deployable gradient-boosted tier classifier under 5-fold cross-validation (93.1% accuracy); off-diagonal errors fall only between neighbouring tiers.

Table 12. Adaptive tiers: size, baseline, post-training, and improvement (offline demonstration on study data, $n = 130$).

Tier	n	Mean Baseline	Mean Post	Mean Improvement
Foundational	46	2.43	4.10	+1.79
Intermediate	60	3.73	4.49	+0.79
Applied	24	4.87	5.02	+0.30

One-way ANOVA of improvement across tiers: $F = 68.8$, $p < 0.001$. Scores on the pooled 1–6 scale.

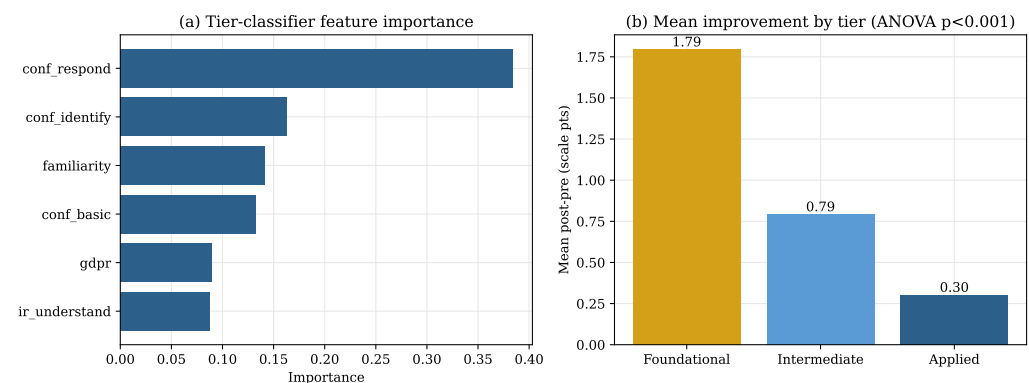


Figure 8. (a) Gini feature importance of the tier classifier; incident response confidence and familiarity dominate. (b) Mean absolute pre–post improvement by tier, showing the gap-closing gradient (Foundational $\gg$ Applied) that validates the tiering as outcome-relevant.

5.3.2. Awareness–Behaviour Gap Detection (Component ii)

Awareness training fails when knowledge does not convert into action. Using only the structured behavioural intention items, the layer computed a behavioural intention score (0–100) for each participant and a post-training competency score, then flagged the “knows but won’t act” segment consisting of participants who were above the median in competency yet below the median in intention. Competency and intention were only moderately correlated ($r = 0.48$), confirming that intention is not redundant with knowledge and must be modelled separately. Nineteen participants (14.6%) fell in the gap quadrant (Figure 9a). Notably, the gap was not concentrated in the lowest-competency tier; it peaked in the

Intermediate tier (23.3% flagged) and was rare in the Foundational tier (4.3%; Figure 9b), indicating that mid-competency participants confident enough to disengage but not yet committed to behavioural change are the segment that a purely knowledge-based program would most easily overlook.

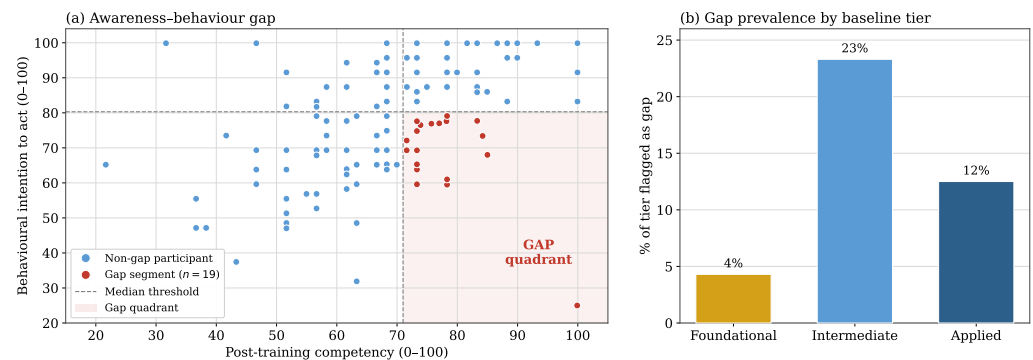


Figure 9. Component (ii), awareness-behaviour gap detection from structured behavioural-intention data. (a) Post-training competency versus behavioural intention; the red points (high knowledge with low intention) are the gap segment ($n = 19$). (b) Gap prevalence by baseline tier, peaking in the Intermediate tier. Both panels use only structured survey fields.

5.3.3. Predictive Early Warning (Component iii)

Detecting the gap at exit is useful; predicting it at entry is more useful. A gradient-boosted classifier was trained to predict the gap flag from baseline structured signals only (baseline competency items, prior security habits, and age) using 5-fold cross-validation. The model identified at-risk participants with 81.5% accuracy and an ROC-AUC of 0.73 (Figure 10a), which is well above chance, with age, baseline confidence, and prior password update habits forming the dominant predictors (Figure 10b). This shows that the layer can pre-emptively allocate behavioural support from the same intake instrument that drives tiering, rather than waiting until a participant has already disengaged.

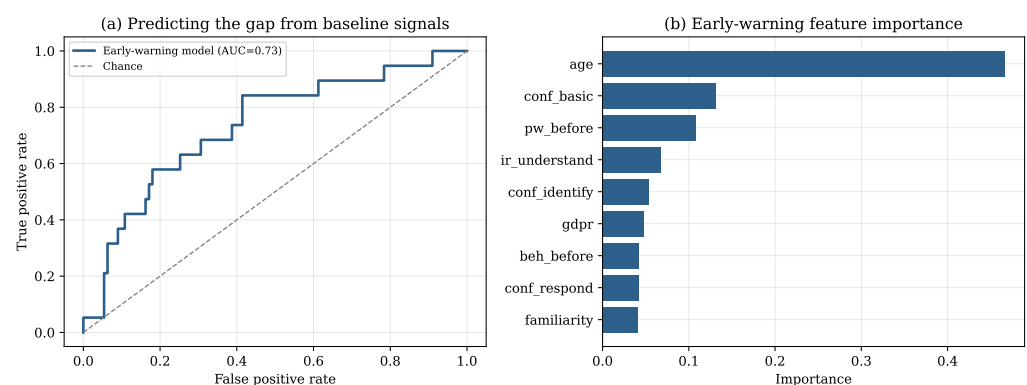


Figure 10. Component (iii), predictive early-warning: (a) ROC curve for predicting the awareness-behaviour gap from baseline structured signals alone ($AUC = 0.73$) and (b) feature importance, where age, baseline confidence, and prior password habits dominate. The model uses only intake structured fields.

5.3.4. Follow-On Recommendation (Component iv)

A transparent rank-tertile rule over the post-training competency vector, with a reinforcement override for any participant retaining a weak domain, was used to assign 45 participants (34.6%) to the Reinforcement track, 42 (32.3%) to the Consolidation track, and 43 (33.1%) to the Advanced track (Figure 11). The distribution aligns coherently with the baseline tiers and with age: reinforcement is concentrated in the Foundational tier and

among older participants, who retained the largest residual headroom, while advanced placement is concentrated in the Applied tier. Therefore, the recommender automatically operationalises demographic responsiveness from each participant's own exit profile rather than from a group stereotype.

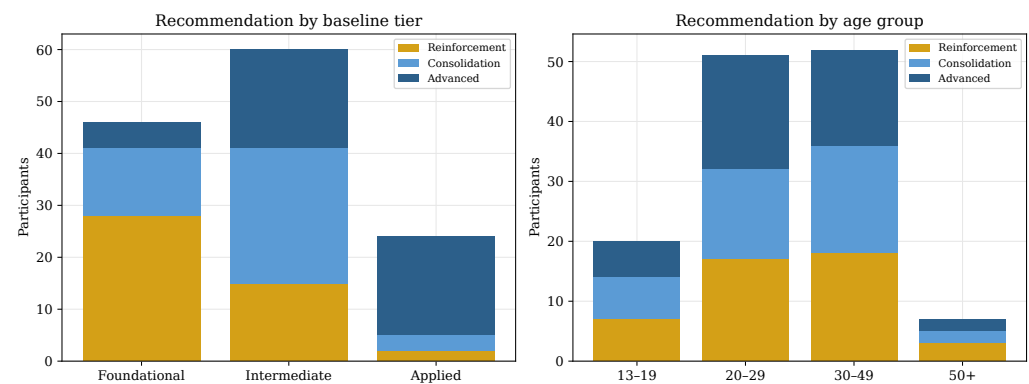


Figure 11. Component (iv), follow-on recommendation: distribution of recommended next-step tracks by (left) baseline tier and (right) age group, computed from structured post-training competency fields.

Together, these four demonstrations show that the AI adaptive layer is implementable on exactly the structured data Cyber4Me already collects, with no reliance on free text: tiering, gap detection, early warning prediction, and recommendation all run directly on the participant vectors with strong internal validity. What they do not show and do not claim is that adding the layer increases learning beyond the embodied sequence, which remains the central question for the controlled study.

6. Discussion

The results reveal a more nuanced picture of cybersecurity training effectiveness than is typically captured in single-modality interventions. Pre-training measures were significantly correlated with each other ($r = 0.39\text{--}0.51$), indicating that the four competency domains are related but not identical. Education was associated with higher baseline GDPR knowledge ($M = 3.09$ for postgraduates vs. $M = 2.75$ for high school or below), consistent with the view that formal education exposes participants to regulatory concepts earlier. Age was not strongly associated with baseline differences.

All four competency domains showed significant pre-post increases (paired- t , $p < 0.001$; $d = 1.03\text{--}1.20$). These findings are not characterised as “causal” effects: the absence of a control group means that some portion of the observed change could plausibly reflect testing effects, demand characteristics, or maturation. What can be said is that the observed within-participant changes are large by conventional Cohen criteria and consistent across demographic subgroups.

Demographic moderation was substantial. Older adults (50+) showed the largest familiarity gain (88.89%), a finding that contradicts the common expectation that older participants gain less from technology training and that is consistent with recent evidence that older adults' engagement is shaped more by delivery format and social context than by capability [23]. Undergraduates showed larger confidence gains than postgraduates, consistent with a ceiling effect (postgraduate baseline confidence was higher, leaving less room for measured improvement). Teenagers showed the largest GDPR-knowledge gain ($n = 5$, so should be treated as preliminary).

These results have three practical implications. First, multi-stage embodied learning programs may be effective for engaging older demographic groups that purely digital tools

can struggle to reach. Second, regulatory training, particularly GDPR-focused content, appears especially beneficial for participants with lower baseline knowledge. This suggests that compliance-oriented material should be introduced early in community-facing cybersecurity education programs. Third, postgraduate participants, who may already possess stronger foundational knowledge and may in some contexts act as designers, facilitators, or trainers, are likely to benefit more from advanced or specialised follow-on content than from introductory awareness material.

The within-participant effect sizes observed in this study are larger than the moderate effects reported in the Prümmer et al. meta-analysis [5] for knowledge and threat recognition outcomes. However, this difference should be interpreted cautiously for three reasons: (i) the study design does not include a control group, which may inflate apparent effects; (ii) the outcomes are self-reported on a 7-point scale, which differs from many studies included in the meta-analytic pool; and (iii) the participant population is community-recruited rather than employee-recruited. Therefore, this study does not claim that Cyber4Me outperforms the reviewed analogs. Rather, the findings indicate that within-participant gains under the Cyber4Me format are substantial and consistent enough to justify future evaluation through a controlled comparative study.

Read against the wider literature, two findings do more than confirm prior work: they qualify it. First, the meta-analytic claim that method modernness and interactivity moderate effectiveness [5] is usually instantiated as “make the digital interface better.” Our results for older adults point the other way: the largest gain in our study came from an embodied low-technology format for the group least served by digital interfaces, which is consistent with Parti et al. [23] that delivery format and social context, not cognitive capacity, gate older adults’ engagement; interactivity, in other words, need not be digital. Second, the acceptance literature [30] predicts that apathy and low trust suppress engagement; the community team-based, voluntary format appears to lower exactly these barriers, which offers a mechanism for why a non-mandatory program produced large gains rather than just an outcome. These points connect the results to theory instead of leaving them as isolated percentages.

The findings also clarify where the integrated AI layer (Section 3.7) adds value, and the offline demonstration (Section 5.3) makes this concrete. The demographic moderation observed here is precisely the signal an adaptive system consumes, and the tiering analysis shows that the signal is strong enough to act on: baseline competency vectors form separable tiers, while a lightweight classifier recovers them at 93.1% accuracy; most importantly, these tiers differ roughly sixfold in measured improvement (Foundational +1.79 vs. Applied +0.30), and an awareness–behaviour gap segment can be detected from behavioural intention data and predicted from intake signals alone. Therefore, a system that starts a Foundational participant on familiarity-building content and an Applied participant on stretch content is responding to a real, large, and predictable gradient, not a stereotype. At the same time, the result that gains were already large without a deployed AI arm is a caution against attributing future improvement to AI by default; the offline demonstration shows that while the layer is implementable and well-targeted, its incremental value over a well-designed embodied sequence is an empirical question that the proposed controlled study must answer, not assume. This is the disciplined reading of the observation that AI is the most consequential feature: it is the feature most worth adding and rigorously testing, not a guaranteed multiplier.

7. Conclusions and Future Work

This study makes three contributions, which we state explicitly rather than restate as numbers. Methodologically, it provides a transparent and demographically stratified pre-

post evaluation of a multi-modal community awareness program, honestly reporting effect sizes, correction procedures, and the boundaries of a single-group design—an evidence type that the AI-enabled analogs in Table 1 largely lack. Empirically, it shows that the demographic pattern of benefit runs counter to common assumptions: the groups usually treated as hardest to reach (older adults, lower-education participants) recorded the largest gains in familiarity and regulatory knowledge, respectively, while higher-baseline groups showed the smallest measured movement, a pattern best explained by a gap-closing and ceiling-bounded interpretation rather than by differential capability. Practically and theoretically it specifies an AI adaptive-layer architecture that operationalises this demographic responsiveness across embodied stages, connecting experiential learning and self-determination theory to a concrete personalisation mechanism. In this way, it reframes the comparison with prior frameworks around fitness-for-purpose rather than feature count.

In practice, the demographic findings translate into concrete design rules: start older adults in familiarity-building scenario-led content; route lower-education participants to regulatory compliance material early on, where their gains are largest; and reserve advanced or facilitator-track content for higher-baseline participants who hit measurement ceilings on introductory material. Each future direction is tied directly to a limitation identified above. First, because this study was short-term (limitation: no follow-up); longitudinal evaluation at 3, 6, and 12 months is needed to test whether the gains persist and translate into behaviour. Second, the design is single-group and self-reported (limitation: no causal control, response bias); thus, a controlled study with a no-training control arm, an active digital-only comparison arm (e.g., a KnowBe4-style intervention), and behavioural or objective-knowledge endpoints is required to estimate efficacy. Third, and most important given the role of AI established here because the AI adaptive layer (Section 3.7) is integrated at the design and pilot level but not yet evaluated as a separate arm (limitation: AI contribution unmeasured), the controlled study should include an AI-on versus AI-off contrast so that the incremental value of adaptivity over the embodied sequence can be measured rather than assumed. Fourth, because the sample is single-region and unevenly stratified (limitation: representativeness), replication across multiple regions and larger teenager and postgraduate cohorts is needed. Together, these steps would convert the present within-participant evidence into a causal, generalisable, and AI-quantified account.

As cybersecurity threats continue to evolve with sophistication, the human element remains both the greatest vulnerability and the strongest defence. By acknowledging and adapting to the demographic diversity within communities, cybersecurity training can be transformed from a perfunctory compliance exercise into a powerful tool for genuine behavioural change.

Author Contributions: Conceptualization, M.I. & M.A.R.; methodology, M.I. and T.S.W.; software, M.I.; validation, M.I., M.A.R., P.P.; formal analysis, M.I.; investigation, M.I. and M.A.R.; resources, B.A.; data curation, M.I.; writing—original draft preparation, M.I.; writing—review and editing, M.I. & P.P.; visualization, M.I.; supervision, M.A.R. & N.R.; project administration, P.P.; funding acquisition, M.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by Department for Science, Innovation and Technology (DSIT) and the University of Wolverhampton as part of the Cyber4Me and Cyber4Me+ projects.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University of Wolverhampton Research Ethics Committee (Approval reference: REC number MCSEC/202526/AR/005.) prior to data collection.

Informed Consent Statement: Written informed consent was obtained from all participants involved in the study.

Data Availability Statement: Data are available upon request. The anonymised data supporting the findings of this study are available from the corresponding author on reasonable request, subject to participant consent restrictions.

Acknowledgments: We thank the anonymous reviewers for their detailed and constructive comments, which substantially improved the manuscript’s framing, methodology, internal consistency, and comparative positioning.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANOVA	Analysis of Variance
APPEARS	Awareness, Perception, Practice, Experience, Achievement, Reward, Socialisation
CAT-RWE	Cybersecurity Awareness and Training framework for Remote Working Employees
CSAT	Cyber-Security Awareness Training
CTF	Capture the Flag
GDPR	General Data Protection Regulation
GPT-CSAT	Generative Pretrained Transformer-based CSAT
HAIS-Q	Human Aspects of Information Security Questionnaire
iCAT	Integrated Cybersecurity Awareness Training
IQR	Interquartile Range
ISO	International Organization for Standardization
MICE	Multiple Imputation by Chained Equations
ML	Machine Learning
NLP	Natural Language Processing
SDT	Self-Determination Theory
SPSS	Statistical Package for the Social Sciences

References

- Chen, S.; Hao, M.; Ding, F.; Jiang, D.; Dong, J.; Zhang, S.; Guo, Q.; Gao, C. Exploring the global geography of cybercrime and its driving forces. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 71. [CrossRef] [PubMed]
- Blythe, J.M.; Johnson, S.D.; Manning, M. The role of human error in successful cyberattacks: A systematic review. *J. Cybersecur.* **2023**, *9*, tyad002. [CrossRef]
- Parsons, K.; Calic, D.; Pattinson, M.; Butavicius, M. Measuring the impact of cybersecurity awareness training on employee behavior. *Comput. Secur.* **2023**, *125*, 102960. [CrossRef]
- Alshaikh, M.; Maynard, S.B.; Ahmad, A. Cybersecurity challenges in small and medium enterprises: A global assessment. *Int. J. Inf. Manag.* **2023**, *68*, 102537. [CrossRef]
- Prümmer, J.; van Steen, T.; van den Berg, B. Assessing the effect of cybersecurity training on end-users: A meta-analysis. *Comput. Secur.* **2024**, *150*, 104206. [CrossRef]
- Ibrahim, M.; Safa, N.S. Detecting message spoofing attacks on smart vehicles. *Comput. Fraud Secur.* **2023**, *2023*, 12. [CrossRef]
- Ibrahim, M.; Safa, N.S. Vulnerabilities and risk in smart vehicle automatic parking assist systems. *Netw. Secur.* **2023**, *2023*, 6. [CrossRef]
- KnowBe4. 2025 *Phishing by Industry Benchmarking Report*; KnowBe4: Clearwater, FL, USA, 2025. Available online: <https://www.knowbe4.com/resources/reports/phishing-by-industry-benchmarking-report> (accessed on 1 April 2026).
- Proofpoint. *State of the Phish 2025: An In-Depth Exploration of User Awareness, Vulnerability, and Resilience*; Proofpoint: Sunnyvale, CA, USA, 2025. Available online: <https://www.proofpoint.com/us/resources/threat-reports/state-of-phish> (accessed on 1 April 2026).
- Taherdoost, H. Towards an innovative model for cybersecurity awareness training. *Information* **2024**, *15*, 512. [CrossRef]
- Hijji, M.; Alam, G. Cybersecurity Awareness and Training (CAT) framework for remote working employees. *Sensors* **2022**, *22*, 8663. [CrossRef] [PubMed]
- Al-Dhamari, N.; Clarke, N. GPT-enabled cybersecurity training: A tailored approach for effective awareness. *arXiv* **2024**, arXiv:2405.04138. [CrossRef]

13. Donekal Chandrashekar, N.; Lee, A.; Azab, M.; Gracanin, D. Understanding user behavior for enhancing cybersecurity training with immersive gamified platforms. *Information* **2024**, *15*, 814. [CrossRef]
14. Birajdar, A.; Nisha, T.N. APPEARS Framework for evaluating Gamified Cyber Security Awareness Training. In *Proceedings of the 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*; IEEE: New York, NY, USA, 2022; pp. 1–8. Available online: https://ui.adsabs.harvard.edu/link_gateway/2022icsi.conf..106B/doi:10.1109/IC3SIS54991.2022.9885399 (accessed on 30 May 2026).
15. Williams, T. Design of a virtual cybersecurity escape room. In *Proceedings of the 2021 IACIS Conference*; Dakota State University: Madison, SD, USA, 2021. Available online: <https://scholar.dsu.edu/bispapers/369> (accessed on 1 April 2026).
16. Keller, T.; Guggemos, J.; Warwas, J. Digital educational escape rooms as a novel approach to cybersecurity education: An empirical study on learner perceptions of usefulness and usability. *Comput. Hum. Behav. Rep.* **2025**, *20*, 100785. [CrossRef]
17. Hadlington, L.; Popovac, M.; Janicke, H.; Yevseyeva, I.; Jones, K. Age, education, and cybersecurity behaviors: A demographic analysis of risk factors. *Cyberpsychol. Behav. Soc. Netw.* **2022**, *25*, 254–262.
18. Hadlington, L. Human factors in cybersecurity: Examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon* **2017**, *3*, e00346. [CrossRef] [PubMed]
19. Vishwanath, A.; Harrison, B.; Ng, Y.J. Suspicion, cognition, and automaticity model of phishing susceptibility. *Commun. Res.* **2018**, *45*, 1146–1166. [CrossRef]
20. Czaja, S.J.; Boot, W.R.; Charness, N.; Rogers, W.A. Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychol. Aging* **2019**, *34*, 341–347. [CrossRef] [PubMed]
21. Salthouse, T.A. Consequences of age-related cognitive declines. *Annu. Rev. Psychol.* **2012**, *63*, 201–226. [CrossRef] [PubMed]
22. Ng, B.Y.; Rahim, M. Adaptive cybersecurity awareness training: A conceptual framework. *Comput. Secur.* **2021**, *108*, 102211. [CrossRef]
23. Parti, K.; Abdelhamid, S.; Ladancsik, T. Designing for Life: A Socioeconomic View of Digital Learning Preferences in Cybersecurity, with Emphasis on Older Adults. *Societies* **2025**, *15*, 342. [CrossRef]
24. Parsons, K.; Butavicius, M.; Delfabbro, P.; Lillie, M. Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q). *Comput. Secur.* **2019**, *80*, 36–44. [CrossRef]
25. Hadlington, L.; Parsons, K.; Furtner, M. Cybersecurity training and awareness programs: A case study of the UK financial sector. *Comput. Secur.* **2021**, *109*, 102352. [CrossRef]
26. Flores, W.R.; Holm, H.; Nohlberg, M.; Ekstedt, M. Investigating personal determinants of phishing and the effect of national culture. *Comput. Secur.* **2020**, *97*, 101823. [CrossRef]
27. Haleem, A.; Javaid, M.; Qadri, M.A.; Suman, R. Understanding the role of digital technologies in education: A review. *Sustain. Oper. Comput.* **2021**, *2*, 129–131. [CrossRef]
28. Flores, W.R.; Ekstedt, M. Shaping intention to resist social engineering through transformational leadership, information security culture and awareness. *Comput. Secur.* **2016**, *59*, 26–44. [CrossRef]
29. Tatum, M.D. Gamification of Security Awareness Training Programs: A Literature Review. Doctoral Dissertation, Middle Georgia State University: Macon, GA, USA, 2023. Available online: https://comp.mga.edu/static/media/doctoralpapers/2023_Tatum_0516152056.pdf (accessed on 1 April 2026).
30. Fallatah, W.; Kävrestad, J.; Furnell, S. Establishing a Model for the User Acceptance of Cybersecurity Training. *Future Internet* **2024**, *16*, 294. [CrossRef]
31. Air & Space Forces Association. *CyberPatriot: The National Youth Cyber Education Program*; AFA: Arlington, VA, USA, 2024. Available online: <https://www.uscyberpatriot.org> (accessed on 1 April 2026).
32. Švábenský, V.; Vykopal, J.; Cermak, M.; Laštovička, M. Cybersecurity knowledge and skills taught in capture the flag challenges. *Comput. Secur.* **2021**, *102*, 102154. [CrossRef]
33. Nicholson, S. Peeking Behind the Locked Door: A Survey of Escape Room Facilities. *White Paper*. **2015**. Available online: <http://scottnicholson.com/pubs/erfacwhite.pdf> (accessed on 1 April 2026).
34. Hasan, K.F.; Hughes, W.; Rahman, A. Gamifying Cyber Governance: A Virtual Escape Room to Transform Cybersecurity Policy Education. *arXiv* **2024**, arXiv:2601.10852. [CrossRef]
35. Srivatanakul, T. Designing cybersecurity escape rooms: A gamified approach to undergraduate learning. *J. Cybersec. Educ. Res. Pract.* **2024**, *2024*, 25. Available online: <https://files.eric.ed.gov/fulltext/EJ1463143.pdf> (accessed on 30 May 2026). [CrossRef]
36. Kolb, D.A. *Experiential Learning: Experience as the Source of Learning and Development*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1984.
37. Deci, E.L.; Ryan, R.M. The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inq.* **2000**, *11*, 227–268. [CrossRef]
38. Faul, F.; Erdfelder, E.; Lang, A.G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [CrossRef] [PubMed]

39. Bognár, L.; Bottyán, L. Evaluating Online Security Behavior: Development and Validation of a Personal Cybersecurity Awareness Scale for University Students. *Educ. Sci.* **2024**, *14*, 588. [[CrossRef](#)]
40. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.