DOI: 10.37943/YXUR4896

Raikhan Muratkhan
PhD, Associate Professor, Department of Applied Mathematics and Informatics
Muratkhan.R@karnu-buketov.edu.kz; orcid.org/0000-0002-2030-8948
Buketov Karaganda National Research University, Kazakhstan
Yedige Radolda (corresponding author)
PhD Student, Institute of Digital Sciences and Artificial Intelligence
radolda_ye_1@enu.kz; orcid.org/0009-0001-9374-0336
L.N. Gumilyov Eurasian National University, Kazakhstan
Darkhan Kydyrali
Master's Degree, Lecturer, Department of Applied Mathematics and Informatics
dar_king1607@mail.ru; orcid.org/0009-0004-2379-2191
Buketov Karaganda National Research University, Kazakhstan
Meiramkul Nurmashova
Master's Degree, Lecturer, Department of Applied Mathematics and Informatics
mnurmashova98@mail.ru; orcid.org/0009-0005-1584-0891
Buketov Karaganda National Research University, Kazakhstan

AN ADAPTIVE LEARNING METHODOLOGY FOR CYBERSECURITY CULTURE THROUGH SIMULATIONS AND GAMIFICATION

Abstract: Most security incidents begin with a person, not a machine. It takes a single rushed employee who trusts a well-made message; no technical control prevents that. The usual organisational answer, a periodic lecture with a test, satisfies the regulator yet hardly moves behaviour, and it hands a new clerk the same course as a security engineer. Kazakhstan feels the mismatch harder, since state and banking services have moved to mobile while security education in Kazakh remains scarce. Here we propose an adaptive-learning methodology for building a cybersecurity culture across user levels. Four elements that the literature keeps apart are joined in it: diagnostic risk profiling, an adaptive learning trajectory, interactive attack simulations under safety constraints, and gamification. A PRISMA-guided bibliometric mapping of 129 Web of Science and Scopus works documents the integration gap. The methodology is written out as a technical specification and realised in a live trilingual web-and-mobile platform (Kazakh, English and Russian) whose behaviourally scored simulation logs feed each learner's risk profile, which in turn decides how the trajectory adapts. Three algorithms do this work: action scoring, risk-profile update, weak-spot course recommendation. Adaptation is rule-based: deterministic and auditable, not learned. Measured live, each finishes server-side within three milliseconds; authenticated endpoints answer in about 60–85 ms end-to-end. A demonstration cohort exercised the path in full: 61 learner accounts, 124 course enrolments (34 completed), 264 graded quiz attempts, mean score 76.7%. That shows feasibility; efficacy is for a randomised controlled trial to show. We contribute a reproducible, culture-oriented, Kazakh-inclusive methodology, its specification, and a working platform on which awareness becomes measurable, adaptive behaviour.

Keywords: risk profiling, cybersecurity culture, security awareness, phishing simulation, gamification, adaptive learning platform, instructional design, Kazakh language.

1. Introduction

Information security is no longer a purely technical discipline. When breaches are examined, a human element turns up in most of them, since social-engineering attacks like phishing work by manipulating the person rather than by defeating the control [1-2]. Organisations usually answer with annual compliance training built around a slide deck and a knowledge test,

Copyright © 2026, Authors. This is an open access article under the Creative Commons CC BY-NC-ND license
Received: 04.07.2026 Accepted: 17.09.2026 Published: 30.09.2026

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

which satisfies the regulator while changing little in what people actually do: the knowledge fades within weeks, and recognising a live phishing message under time pressure draws on a different skill than knowing, in the abstract, that phishing exists [3-4]. There is a second weakness, too. Most programmes treat every user alike, even though susceptibility varies with age, role, experience and language [5].

The problem sharpens in Kazakhstan and Central Asia. Services keep moving to phones, attack volumes keep growing, and the research base lags the languages involved: the datasets, tools and training content used against phishing exist mostly in English, partly in Russian, and in Kazakh hardly at all [6-7]. Nor are Kazakh-language lures rare any more — messages styled as familiar government-service or parcel notifications reach users no one has taught to distrust them. A person who has never once seen an attack in their own language simply has no mental model for spotting one.

What follows from both observations is not a call for more slides. One-size-fits-all awareness has had its chance. We argue instead for an adaptive, behaviour-based approach that does four things at once: it measures each user's actual cyber-behaviour, matches the learning path to their level, provides realistic yet safe practice in interactive attack simulations, and keeps people engaged through gamification, all delivered natively in Kazakh, English and Russian. The article develops this approach as the conceptual and design foundation of a cyber-literacy platform and then pins it down concretely: a technical specification, gamification mechanics, UX/UI recommendations.

The argument runs as follows: Section 2 reviews related work and, through a PRISMA-guided bibliometric mapping, establishes the integration gap this work fills, after which Section 3 states the aim and objectives and Section 4 sets out the materials and methods behind both the methodology design and the bibliometric procedure. The results come in Section 5 — the methodology, its technical specification and a platform realisation — while Section 6 weighs implications and limitations before Section 7 concludes.

2. Related work and problem statement

2.1. Conceptual background

The methodology stands on three bodies of work. From behavioural security we take the finding that unsafe clicks are mostly a product of habitual, low-attention processing [8], and that risk concentrates: a small group of repeat clickers accounts for a disproportionate share of incidents [5, 9]. Awareness training and simulation research supplies the second. Embedded, repeated phishing drills hold up longer than lectures do, and the gain traces mainly to spaced practice [10] and to immediate feedback [3-4]; a scoping review adds the caveat that susceptibility shrinks but never quite disappears [11]. Adaptive learning and gamification supply the third: personalization and well-made game mechanics go hand in hand with engagement and self-efficacy [12-13], an association that systematic reviews of gamified security education keep finding [14]. Rarely, though, do the three literatures meet, and uniting them in a single culture-forming methodology — for a low-resource language at that — is precisely the gap this work goes after.

2.2. Systematic mapping procedure (PRISMA-guided)

To make the gap explicit and reproducible, the mapping follows a systematic mapping study (SMS) protocol [15-16], reports according to the PRISMA guidelines [17], and draws on established bibliometric methods [18] — the same protocol that has already been applied in related Kazakhstani work [19].

Search structure (PICOC). The population comprises users and employees receiving cybersecurity awareness or education; the intervention covers adaptive or personalized learning, interactive phishing and attack simulation, and gamification; no specific comparator is set; the outcomes of interest are behaviour change, susceptibility, engagement and cybersecurity culture; and the context is academic studies of security education and training.

Search string (Scopus, run 1 July 2026): TITLE-ABS-KEY(("security awareness" OR "cybersecurity awareness" OR "cyber security awareness" OR "information security awareness"

DOI: 10.37943/YXUR4896
© Raikhan Muratkhan, Yedige Radolda,
Darkhan Kydyrali, Meiramkul Nurmashova

OR “security culture” OR “cyber security culture” OR “cyber hygiene” OR “cybersecurity education” OR “cyber security training”) AND (“adaptive learning” OR “personalized learning” OR “personalised learning” OR gamif* OR “serious games” OR “game-based learning” OR “phishing simulation” OR “simulated phishing” OR “simulation-based” OR “embedded training”)) AND PUBYEAR > 2014 AND (DOCTYPE(ar) OR DOCTYPE(re)) AND LANGUAGE(english) AND (SUBJAREA(comp) OR SUBJAREA(soci) OR SUBJAREA(engi)). An equivalent Topic query (field tag TS) was run in the Web of Science Core Collection over the same period, restricted to articles and reviews in English. The two queries and their yields are summarised in Table 1.

Table 1. Search queries and yield.

Database	Document type	Language	Subject areas	Years	Results
Web of Science (run 2 July 2026)	article OR review	English	Computer Science · Education · Information Science	2015–2026	98
Scopus (run 1 July 2026)	article OR review	English	COMP · SOCI · ENGI	2015–2026	117
Total (two sources; 86 duplicates merged)	—	—	—	—	215 → 129 unique

Selection. Every record went into a reference manager (Mendeley/Zotero), duplicates were removed, and two reviewers then screened the titles and abstracts — keeping a study when it addressed cybersecurity awareness or education and involved adaptivity, simulation or gamification, and dropping purely technical detection papers without a learning or behavioural component along with off-topic keyword matches and superseded duplicates. Of what the two databases returned, 98 records from Web of Science and 117 from Scopus, merging and the removal of 86 cross-database duplicates left 129 unique works for the mapping (Figure 1); reporting both sources together with the exact query strings keeps the procedure reproducible, in line with established two-source bibliometric mappings.

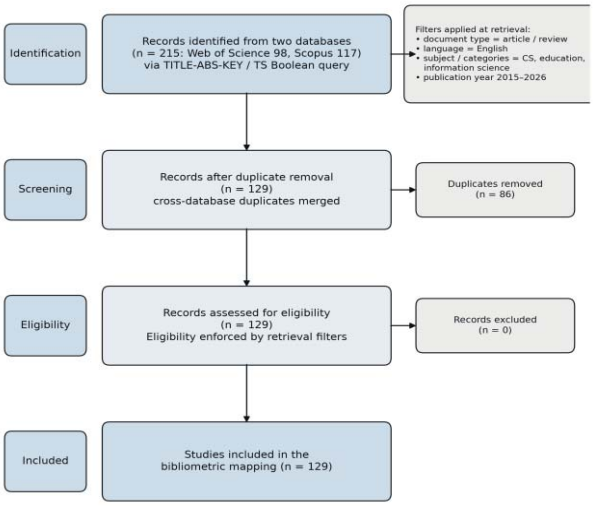

Figure 1. PRISMA flow of the two-source (Web of Science and Scopus) systematic mapping (n = 129).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

2.3. Bibliometric analysis (VOSviewer)

Term co-occurrence over the 129 retrieved works (VOSviewer [20]; Web of Science and Scopus author and index keywords, full counting, thesaurus-cleaned; minimum-occurrence threshold 5) produced 38 terms grouped into four clusters (Figure 2, a; Table 2). The network is dominated by three hubs — cybersecurity (80 occurrences), security awareness (48) and gamification (45) — with cybersecurity education (39) and serious games (37) as secondary centres, visible as the peaks of the density map (Figure 2, b). The four clusters read as follows:

- (C1) Cybersecurity education and learning technology — the core cluster: cybersecurity (80), cybersecurity education (39), game-based learning (16), network security (14), education computing (12), learning systems (9), personnel training (9), computer-aided instruction (8), federated learning (8), teaching (8), artificial intelligence (8), e-learning (6), cyber threats (5), student engagement (5), capture the flag (5).
- (C2) Security awareness, phishing and human factors — security awareness (48), training (20), phishing (18), computer crime (14), social engineering (10), motivation (10), behavioural research (8), engineering education (7), human–computer interaction (7), computer games (6), human factors in cybersecurity (6).
- (C3) Gamification and serious games — gamification (45), serious games (37), education (20), design (13), impact (8), performance (7), behaviour (7), intention (5). Gamified security learning is an established pairing.
- (C4) Security training and the review base — security training (10), security of data (8), systematic literature review (8), game (5).

Colouring the same map by mean publication year, 2020–2025, gives the temporal overlay of Figure 2, c. Its clearest signal concerns recency: human-factor and artificial-intelligence framings entered the corpus only lately.

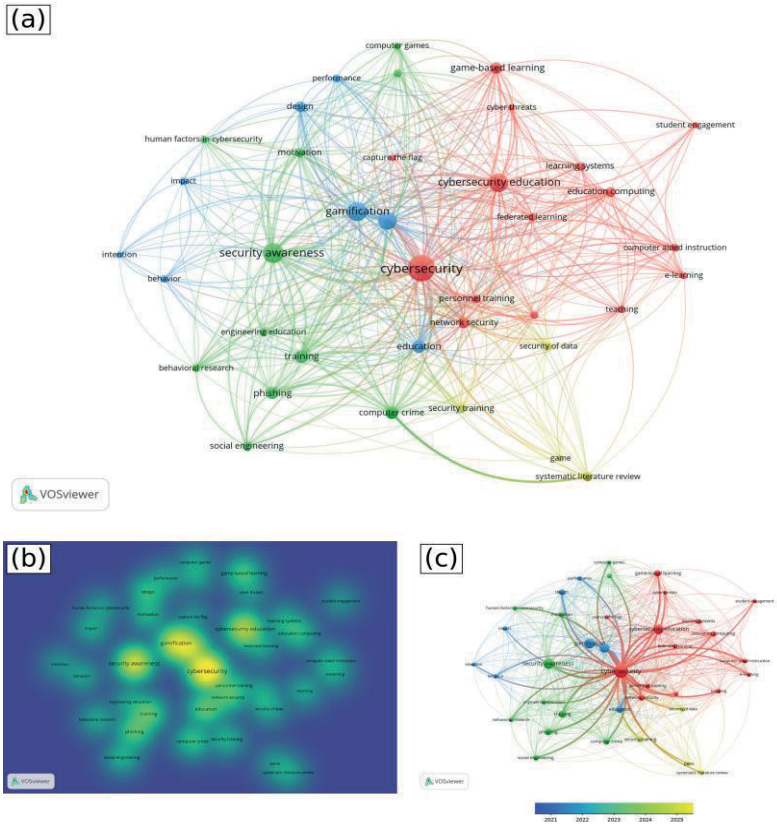

Figure 2. Keyword mapping of the 129-work corpus: a – co-occurrence network (38 terms in four clusters); b – density map (the peaks mark the most frequent terms); c – temporal overlay coloured by average publication year (2020–2025).

Table 2. Terms per cluster (VOSviewer, thesaurus-cleaned map; occurrences in parentheses)

Cluster (theme)	n	Leading terms (occurrences)
C1 — Cyber-security education and learning technology (core)	15	cybersecurity (80), cybersecurity education (39), game-based learning (16), network security (14), education computing (12), learning systems (9), personnel training (9), computer-aided instruction (8), federated learning (8), teaching (8), artificial intelligence (8), e-learning (6), cyber threats (5), student engagement (5), capture the flag (5)
C2 — Security awareness, phishing and human factors	11	security awareness (48), training (20), phishing (18), computer crime (14), social engineering (10), motivation (10), behavioural research (8), engineering education (7), human–computer interaction (7), computer games (6), human factors in cybersecurity (6)
C3 — Gamification and serious games	8	gamification (45), serious games (37), education (20), design (13), impact (8), performance (7), behaviour (7), intention (5)
C4 — Security training and the review base	4	security training (10), security of data (8), systematic literature review (8), game (5)

Reading the map against the study's three levers, gamification (C1/C4/C2) is firmly bound to awareness, while simulation and personalization are not — a reading that Section 2.4 formalises.

2.4. Problem statement (from the mapping)

What does the map say about the gap? Three things. The levers this study cares about sit in different places: gamification and serious games (cluster C3) and game-based learning (C1) keep company with security awareness (C2), while attack simulation shows up only under the term phishing, inside the awareness-and-human-factors cluster (C2) and nowhere near the game mechanics; not a single term ties those mechanics to adaptive, action-scored simulation. Adaptivity hardly figures at all. One of the 129 works uses adaptive learning, two use personalized learning; both counts fall below the occurrence threshold, so no node forms, personalization links to neither gamification nor simulation, and no security-culture metric ever appears. Then there is language. The map contains no Kazakh, multilingual or low-resource-language term, no included work has a Kazakhstani affiliation, and English dominates the corpus from end to end. Judging by the temporal overlay (Figure 2, c), human-factor and AI-supported framings arrived only in 2024–2025, and adaptive, culture-forming, multilingual training is still ahead. From all of this the problem follows: no validated methodology yet joins adaptive learning, safe interactive attack simulation and gamification into cybersecurity-culture formation for heterogeneous, multilingual user groups, Kazakh speakers among them.

3. The aim and objectives of the study

Aim. The study sets out to build an adaptive-learning methodology that forms a cybersecurity culture across users of different levels, combines interactive attack simulations with gamification, and is expressed as a technical specification for a digital learning platform.

The objectives are as follows:

1. Map the literature with a PRISMA-guided bibliometric procedure to surface the dominant themes and the integration gap (Section 2).

2. Specify the pedagogical model: diagnostic profiling of the user's level, an adaptive learning trajectory, and behavioural practice spread over time (Section 4).

3. Design the mechanism that generates attack-simulation scenarios inside a safety constraint (Section 4).

4. Set out the gamification mechanics along with a metric for resilience and culture (Sections 4–5).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

5. Draw out UX/UI recommendations for digital content delivered in Kazakh, English and Russian (Section 5).

6. Consolidate the above into a technical specification and illustrate it with a working platform realisation (Section 5).

4. Materials and methods

4.1. Methodological approach

Two strands make up the study. The first is the systematic mapping of the literature described in Sections 2.2–2.3, which grounds the design. The second is a design-science development of the methodology and its technical specification, instantiated and illustrated in a working platform (CyberQadam).

4.2. Adaptive-learning methodology (design)

Four linked components make up the methodology.

(1) Diagnostic profiling. Each user begins with a short mixed assessment that pairs knowledge items with a brief behavioural simulation. From the result the platform builds a risk and resilience profile (an interpretable score with per-competency and per-language sub-scores) and assigns a level (for example Novice → Aware → Resilient → Expert). This is how the methodology handles "users of different levels".

(2) Adaptive trajectory. Courses, modules and simulations are arranged into a learning roadmap governed by prerequisites, from which the platform selects each next step according to the learner's profile and their weakest competencies. Here the word adaptive needs qualifying, since it carries heavier connotations elsewhere in the literature: adaptation in this system is rule-based, meaning that a deterministic and configurable rule layer reads the profile and chooses the next step, with no policy learned from data at any stage. Auditability is what that choice buys. Trace any routing decision backwards and out come the rule and the profile values that produced it, which is precisely what a safety review demands of a training system intended for public deployment. Calling the methodology adaptive therefore means rule-driven personalisation of the trajectory, not reinforcement learning. A learning controller is specified as an extension point (Section 6). It is not claimed as a property of what runs today.

(3) Interactive attack simulation (behavioural practice). Learners work through safe but realistic phishing and social-engineering messages, choosing an action for each: report, treat as safe, delete or open. The system logs the action, whether it was correct, and the reaction time, and it withholds the red flags and the explanation until the learner has committed. Practice is therefore spaced and feedback-rich, and it produces the signals that drive adaptation.

(4) Gamification and culture formation. Engagement rides on experience points, daily streaks, a leaderboard, badges, a visual roadmap and a resilience metric watched over time. The intent behind these mechanics is cultural — security becomes a shared competence that keeps improving, not a test passed once.

4.3. Attack-scenario generation mechanism (safety-constrained)

Scenarios come from parameterised trilingual templates covering a spread of social-engineering themes at three difficulty levels; every template records the cue it plays on (urgency, authority or reward). Nothing that comes out of the generator can be deployed. Three layers see to it. Domains are always placeholders. Real brands and credential-harvesting phrases get blocked at runtime, every item carries a "[SIMULATION – RESEARCH USE ONLY]" watermark, and every record keeps its provenance flags. As for the generator itself, it is swappable: parameterised templates run it today, and large-language-model generation is a defined extension.

4.4. Bibliometric method

The tools and settings follow Sections 2.2–2.3. Records were retrieved from two databases, Web of Science and Scopus, de-duplicated in a reference manager, and analysed in VOSviewer with a thesaurus to produce the co-occurrence, density and temporal-overlay maps. For reproducibility, the dataset (query, records, thesaurus and map) is released in a public repository.

5. Results

5.1. An evidence-based gap

The first result is evidential: the mapping of Section 2. From 129 works it extracts 38 co-occurring terms in four clusters (Table 2; Figure 2), and there the three levers of interest (adaptive or personalized learning, interactive attack simulation, gamification) occupy separate clusters with only weak bridges between them. Adaptive learning does not even clear the occurrence threshold. Kazakh never appears, no multilingual term appears, and no included work is Kazakhstani-affiliated. Even so, the 2024–2025 output drifts toward human-factor and AI-supported framings (Figure 2, c), which is the direction this work takes. A methodology that unifies the three levers, keeps capability language-specific and puts Kazakh at the centre is what such evidence calls for. The rest of Section 5 lays it out.

5.2. The methodology as a technical specification

The methodology of Section 4.2 is set out as a technical specification for a digital learning platform, then realised in a working instance: a public web deployment with native mobile clients. Figure 3 gives the architecture: trilingual (Kazakh/English/Russian) web and mobile clients over a Django REST core. A safety-constrained scenario generator feeds the four methodology engines (diagnostic profiling, adaptive trajectory, interactive simulation and gamification), and a per-learner risk/resilience profile holds the state they share. Details a level lower (role-based access, the course → module → lesson → quiz content model, analytics) belong to the platform's technical documentation and stay there. Each item of content exists in all three languages and resolves to the interface language the user picked. As of July 2026 the deployment is publicly operational, and its use so far paints the picture: three courses are live, carrying 27 lessons and nine weighted quizzes alongside five trilingual simulations, while the 61 learner accounts served — 51% preferring Kazakh, 21% English and 28% Russian — have produced 124 course enrollments, 34 of them complete, and 264 graded quiz attempts whose weighted average comes to 76.7%. Because those figures mix a demonstration cohort with early public accounts, what they establish is not training efficacy but the thing prior to it, namely that every component of the specification runs end-to-end in production (Section 6).

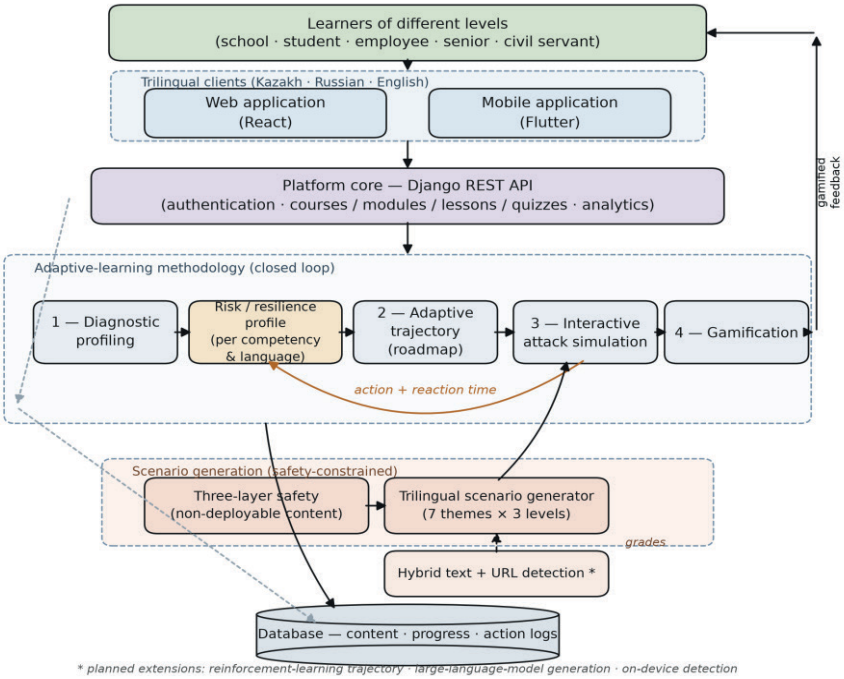

Figure 3. Architecture of the adaptive-learning methodology and its platform realisation; items marked * (reinforcement-learning trajectory, large-language-model generation, on-device detection) are planned extensions.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

5.3. Diagnostic profiling and the capability map (component 1)

Figure 4 shows the diagnostic output. From the learner's assessments and behavioural history, the platform derives a cyber-resilience score on a four-level scale (Novice → Aware → Resilient → Expert), a set of per-language sub-scores, and a list of weak competencies. At its centre sits the capability heatmap: competency × language, cells scored 0–100 and coloured red to green, one view answering where the learner is weak, in which topic and in which language. In the illustrated profile English clearly outruns Kazakh (EN ≈ 82 vs KK ≈ 54); social engineering and malware sag; link inspection is flagged outright as a remediation target. A single global score cannot capture that. The differentiated, language-aware diagnosis is exactly what the methodology needs; it puts the diagnostic-profiling step (component 1) into practice, and that step grounds the adaptive trajectory.

Figure 4. The cyber-resilience diagnostic: resilience score and level, per-language sub-scores, and the competency × language capability heatmap.

5.4. Adaptive trajectory and gamification (components 2 and 4)

The adaptive roadmap appears in Figure 5, rendering the curriculum as a visual journey: course, simulation and milestone nodes wired together by prerequisites, each node showing its state: locked, available, in progress, completed or mastered. Progression is threaded through with experience points and a daily streak, while an overall journey-progress indicator stays visible in the header. Next moves come from a suggestions panel ("Continue…", "Start…"), which the resilience profile also steers toward targeted remediation ("Revisit your weak area: link inspection"). So one artefact, the roadmap, carries the adaptive trajectory (component 2) and gamified engagement (component 4) at the same time. At any moment the learner can see where they stand, what comes next, and why. A flat catalogue turns into a personalised, motivating path.

DOI: 10.37943/YXUR4896
© Raikhan Muratkhan, Yedige Radolda,
Darkhan Kydyrali, Meiramkul Nurmashova

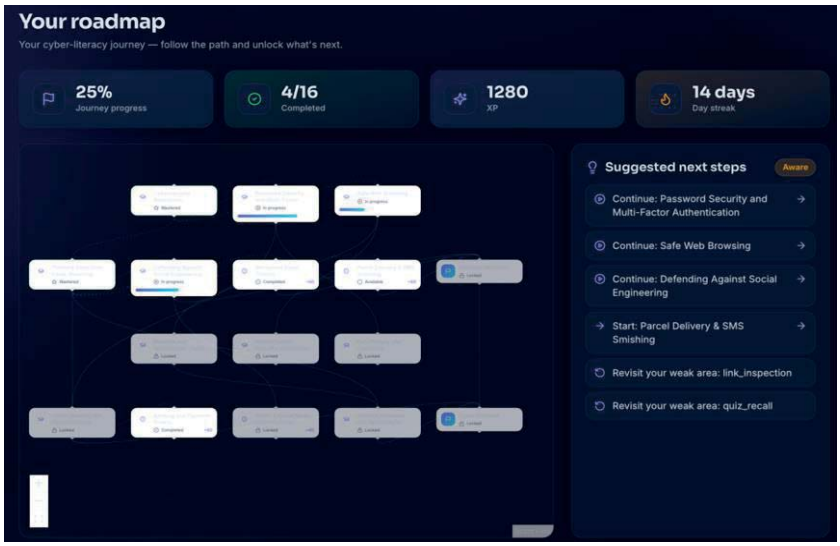

Figure 5. The gamified adaptive roadmap: progression metrics (experience points, streak, journey progress) and profile-driven next-step suggestions.

5.5. Interactive attack simulation (component 3)

The behavioural core is shown in Figure 6. A simulation lays out an inbox of safe messages, and for each one the learner picks an action: Report, Trust, Delete, or Open where a link is present. Each choice is captured — the action taken, whether it was correct, and how long the learner took. Deployability is designed out, since senders and links use fictional placeholder domains such as mega-lottery-prize.example under the three-layer safety architecture (Section 4.3). When the run ends, the learner receives a per-message review (Figure 6, b) that sets the chosen action against the best one, shows the ground-truth label and the red flags, explains in plain language why the message is phishing, and reports an accuracy score together with the average reaction time, 2.5 s in the illustrated run. Every decision therefore becomes spaced, feedback-rich practice and produces the action-and-latency signals that feed back into the resilience profile (Section 5.3) and the roadmap (Section 5.4). That feedback closes the methodology's loop from behaviour to measurement to adaptation.

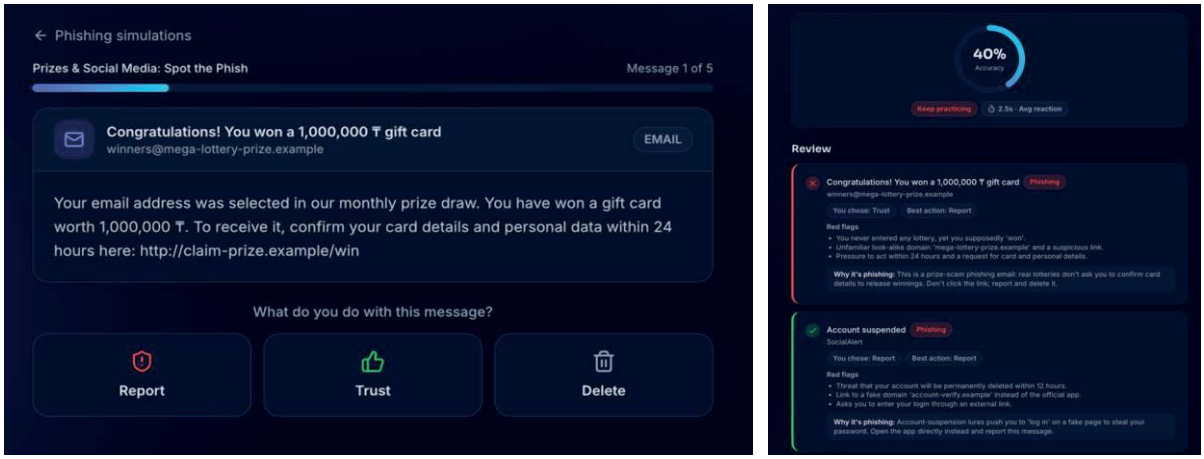

Figure 6. An interactive phishing-inbox simulation: a – a message with the action choices; b – the after-action review with red flags, explanation, accuracy and reaction time.

5.6. Trilingual content and UX/UI recommendations

Figure 7 shows delivered content (a lesson and a scenario-based quiz) rendered in the language the user has chosen. Every course, lesson, quiz and simulation message is served natively

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

in Kazakh, English and Russian, and can be switched at any time. Producing and delivering that content gave us the five UX/UI recommendations the task asks for. Parity across the three languages has to be real: switch the language and the content itself re-resolves, not just the interface labels. Practice sits in message shells everyone already knows, e-mail, SMS and messaging apps, because familiarity is what lets the habit transfer to daily life. Feedback is deliberately withheld until the learner has committed, because red flags and explanations that arrive after the decision keep the assessment valid rather than steering it. A capability view reads better as an interpretable, colour-coded display than as raw numeric scores, and the content itself has to stay readable across mixed Cyrillic and Latin scripts and across a wide span of age groups.

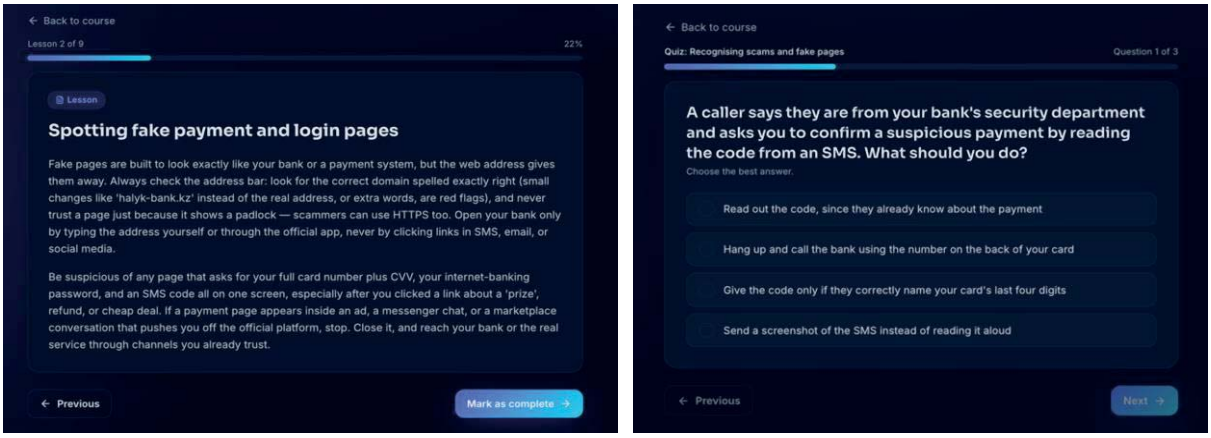

Figure 7. Trilingual learning content: a – a lesson; b – a scenario-based quiz.

5.7. Algorithmic core and measured runtime performance

Underneath the four engines runs a small set of deterministic algorithms that convert each learner action into a score, a risk update and a next step. All of them are rule-based heuristics, auditable by design (every decision can be explained), and they form the substrate the planned reinforcement-learning controller (Section 6) will later replace. Table 3 characterises the four core algorithms by function, computational complexity, output behaviour and measured runtime. A weighted three-tier rule scores each simulation: every chosen action carries a weight in [−1, 1] relative to the message's best action, trusting a phishing message therefore actively costs points, and the run score is the clamped share of the maximum attainable weight. Quizzes are marked question by question, partial credit allowed. Both signals reach the risk-profile update, which nudges a 0–100 risk score, refreshes per-language sub-scores together with the weak-competency set, and lands the result on one of four levels (novice, aware, resilient, expert). If someone underperforms, the weak-spot recommendation looks at competency tags and picks the published course that covers the weakness best, weighted by learner type and leaning toward the easier courses.

Table 3. Core adaptive-learning algorithms: function, complexity, output behaviour and measured runtime.

Algorithm	Function	Complexity	Output / behaviour	Latency, ms (mean / p95)	DB queries
Simulation action-scoring	Weighted three-tier grading of inbox actions against the correct action; logs reaction time	O(m)	0–100 score, pass/fail; action weights in [−1, 1]	1.8 / 2.5	9
Quiz grading	Per-question partial-credit grading; XP award and course-progress recompute	O(q)	0–100 score, pass/fail	2.3 / 2.8	14

Risk-profile update	Heuristic risk nudge; per-language sub-scores; weak-competency set; four-band level	O(1)	risk 0–100; level in {novice, aware, resilient, expert}	0.5 / 0.5	2
Weak-spot recommendation	Competency-overlap course match, user-type weighted, difficulty-biased	O(c)	recommended course + matched competencies + trilingual guidance	0.2 / 0.2	1

Table 3 uses m for messages per simulation, q for questions per quiz and c for published courses; the deployment currently has m = 5, q = 3 and c = 3. The runtimes come from the reference backend (Django 5 / DRF) on a development host, where each algorithm ran 200 timed repetitions after 20 warm-up calls and per-call SQL query counts came from the framework's database instrumentation. Absolute latencies follow the hardware and the database. Complexity classes and query counts do not. All four algorithms stay under three milliseconds and issue a small, bounded number of queries, which leaves the scoring-and-adaptation loop adding next to nothing to request handling while scaling linearly with content size. Production got its own measurement: on the public deployment (cyberqadam.kz, gunicorn behind nginx, PostgreSQL 16), authenticated read endpoints (the learner dashboard, the adaptive roadmap) come back in roughly 60–85 ms on average, end-to-end from an external client, network round-trip and TLS included. The server-side algorithmic cost above is a small fraction of that budget.

5.8. Behaviour under concurrent load

Single-call cost, the subject of Section 5.7, says nothing about what happens when many learners arrive at once, and for a platform proposed for national use that is the question that matters. So the reference backend was driven through a concurrency sweep. Gunicorn ran the production configuration of three synchronous workers against PostgreSQL, and virtual users rose from 1 to 500, each holding its own account and its own token so that the load models distinct simultaneous learners rather than one session replayed. Every level ran 3 s of discarded warm-up followed by 12 s of measurement. Five authenticated read endpoints were exercised, one per architectural component; Table 4 reports the learner dashboard, the heaviest composite read of them, and Figure 8 plots throughput and tail latency for all five.

Throughput climbs steeply and then flattens, which is what a worker-bound service should do. Saturation arrives at roughly 1,199 requests per second on the dashboard, and beyond that point additional concurrency buys queueing rather than work: median latency grows in proportion to the queue while throughput holds, the divergence visible in Figure 8 as flat throughput against a rising tail. Across all five endpoints the error-free ceiling ranged from 336 to 1,199 requests per second, the lower figure belonging to leaderboard. Up to 100 simultaneous users every endpoint answered without a single failed request, the worst 95th percentile at that level being 304 ms. Failures begin at 200 concurrent users and reach 15% of attempts at the top of the sweep, and their character matters more than their count, since each is a refused connection at the accept queue and nothing that reached the application returned a server error. The socket backlog in front of three workers saturates first, which is a provisioning limit rather than a defect.

What does that mean in learners? An active learner issues perhaps one request every 5 seconds, not one every few milliseconds, so 336 requests per second corresponds to roughly 1,680 concurrently active learners per instance, with horizontal scaling available beyond that because no session state is held between requests. Three caveats bound this. Client and server shared one development host, so the figures exclude network and TLS cost and are not end-user latencies, which Section 5.7 reports instead; co-residency also means the load generator competed for the same cores, which if anything understates the ceiling. The local database was PostgreSQL 14 rather than the 16 in production. And per-user rate limiting, which stands at 1000 requests per minute in the deployment and is itself a load control, was lifted for the run, because a closed-loop driver would otherwise have measured the limiter instead of the server.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

Table 4. Learner-dashboard endpoint under rising concurrency: sustained throughput, latency percentiles and failed-request share.

Concurrent users	Throughput, rps	p50, ms	p95, ms	p99, ms	Failed, %
1	301	3.2	3.8	5.4	0
10	1,181	8.3	9.9	11.5	0
25	1,131	21.1	28.6	31.3	0
50	1,175	41.7	49.1	52.8	0
100	1,199	83.1	88.2	93.8	0
200	1,071	110.0	199.0	635.1	0.43
350	1,084	120.4	223.9	694.4	2.26
500	1,030	176.2	304.0	2185.6	3.42

Table 4 reports the dashboard endpoint (/analytics/me/); rps is requests per second. Concurrency is the number of virtual users, each with its own account and JWT. Throughput and percentiles cover the measured window only, warm-up excluded. The failed share counts connections refused once the accept queue saturated; no request that reached the application returned a server error.

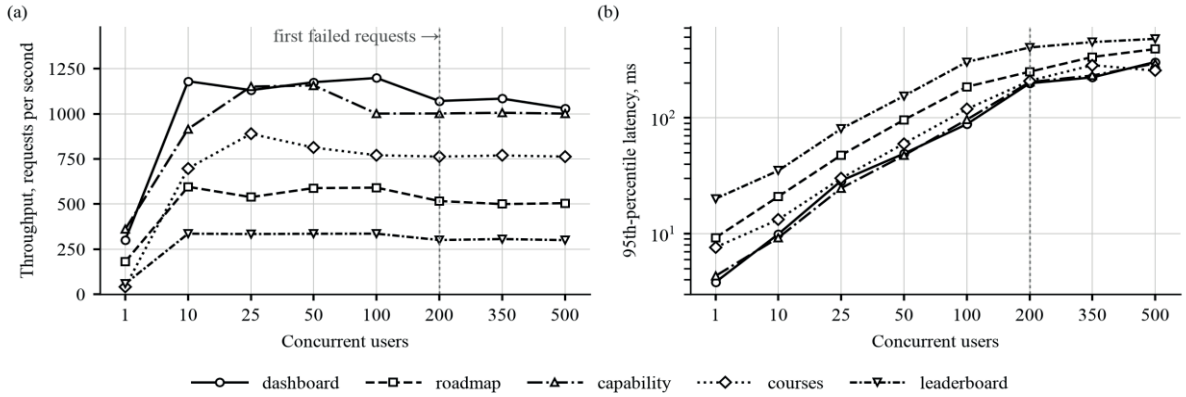

Figure 8. Five authenticated read endpoints under rising concurrency: a – sustained throughput; b – 95th-percentile latency (logarithmic scale). The dashed line marks the concurrency at which the first requests failed.

6. Discussion of results

Three findings came out of the mapping in Section 2: the levers are studied apart, adaptivity is thin on the ground in the awareness literature, and Kazakh, like multilingual settings generally, is close to invisible. The methodology answers all three at once, not with awareness content, simulated attacks and game mechanics bolted together as add-ons, but with one closed loop. A diagnostic sits at the start of the loop and initialises a resilience profile, held per competency and per language. The profile then steers an adaptive roadmap. The roadmap schedules content and action-scored simulations. Whatever the learner does (the action itself, the reaction time) returns to the profile, and gamification keeps the whole thing in motion. Where does this depart from prior work? In two commitments. Capability gets modelled language by language (the Section 5.3 heatmap), so the learner who is strong in Russian and exposed in Kazakh, common enough among Kazakh speakers, is diagnosed and remediated precisely there; one global score would have hidden that exposure. And Kazakh is first-class throughout: generation, content, assessment, interface. That is the gap the mapping surfaced, answered directly.

Each part of that loop rests on established results. Practice is active and experiential, not passive reading [3-4]. It is spaced over time, because the roadmap and the streak mechanic encode the spacing effect [10]. Engagement leans on game mechanics, whose payoff depends on the quality of the content behind them [13, 21]. Assessment records what a learner does under time

pressure, not what they can recall on demand [9-11]. The levers are integrated, not run side by side, and that integration is what drives durable behaviour change for users of different levels, the aim we set out with and the thing one-off uniform training keeps failing to deliver.

The gains land with two audiences in practice. An organisation, to start, learns from the per-language, per-competency diagnosis who is exposed and in which language; and since generation is safe by design, the simulations can run in consented, organisation-scoped settings and cannot yield deployable attack content. The nation is the second audience, and Kazakhstan is the case at hand. Its public and financial services are mobile-first, its Kazakh-language security education is scarce, and a trilingual, Kazakh-inclusive methodology therefore has room to lift cyber-literacy at population scale, one layer above the practitioner-focused academies dominating the local ecosystem.

6.1. Limitations and threats to validity

The claim should be stated precisely. It is a methodology and its working realisation; the effect on behaviour remains unmeasured. What the results establish is coherence and buildability (a deployed trilingual instance embodies each component), and they set the stage for the randomised controlled trial that would quantify durable change. Two engineering choices are provisional, deliberately so. Scenarios come from templates, not from a generative model, and adaptation runs on rules, not on learning. Templates won for a simple reason: each simulated lure must clear the three-layer safety constraint, and with a fixed set every message stays auditable and every domain provably fictional, an assurance free-form generation cannot yet match. The extension points are specified for all of it: large-language-model generation, a reinforcement-learning controller and a direct competency-by-language assessment.

Three gaps in the evidence follow from how that cohort was assembled. Sixty-one accounts are enough to show the loop running end to end and not enough for anything else; split them by age band, by professional role, or by preferred language and the cells collapse into single digits, far below what any stratified comparison would need. There was also no comparison arm. Everyone enrolled used the platform, nobody worked through a conventional briefing alongside them, and so our argument that this approach improves on the lecture-and-test format rests on the published record [3-4] and not on anything measured here. Retention is the third gap and the most consequential of them, because outlasting the session is precisely what a culture has to do. Whether a habit formed inside a simulation is still there after one month, or after six, went unmeasured.

What those gaps call for is a trial, and its design is settled rather than gestured at. Two arms, randomly assigned: conventional periodic training against the adaptive path described here. Action accuracy on a held-out battery of simulated messages serves as the primary outcome, with reaction time and the per-language competency scores secondary to it. Detecting a medium effect of $d = 0.5$ at the conventional five per cent level with 80% power calls for roughly 64 participants per arm; once stratification by language and by role is allowed for, together with attrition across the follow-ups, the recruitment target rises to about 300. Measurement then repeats immediately after training and again at one, three and six months, which is the step that turns a short-term score into evidence about retention. Until that trial reports, what stands is a methodology, the specification behind it, and a working system on which the measurement can actually be made.

7. Conclusion

This article set out to propose, and then fully specify, an adaptive-learning methodology for forming a cybersecurity culture among users of different levels, with interactive attack simulations and gamification working as one mechanism rather than two tools. The case rests on evidence. When 129 works from Web of Science and Scopus went through a PRISMA-guided systematic mapping and VOSviewer analysis, adaptivity, simulation and gamification turned out to be studied in isolation, and Kazakh was missing from the corpus entirely. Four linked components make up the answer: diagnostic profiling around a per-competency, per-language resilience profile; an adaptive learning trajectory; safety-constrained interactive attack simulation that scores the action and the reaction time; and gamification. A scenario-generation mechanism

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 27, SEPTEMBER 2026

under a three-layer safety constraint completes the design, together with UX/UI recommendations for trilingual content in Kazakh, English and Russian. All of this stands in a technical specification and in a publicly deployed working platform illustrating each component (Figures 3–7). Server-side, the algorithmic core stays under three milliseconds; the Section 5.2 demonstration cohort (61 learner accounts, 124 course enrolments, 264 graded quiz attempts, mean score 76.7%) confirms end-to-end operation, which is feasibility rather than efficacy. Under concurrent load the same reference backend answered 100 simultaneous users without a single failed request, saturating between 336 and 1,199 requests per second depending on the endpoint (Section 5.8). Item for item, the outputs match the funded task's completion form — training methodology, technical specification, gamification mechanics, UX/UI recommendations.

The principal contribution, then: an evidence-grounded, reproducible, Kazakh-inclusive methodology that takes three levers long kept separate and folds them into one behaviour-measuring adaptive loop. The limitations write the future-work list themselves. Instantiate reinforcement-learning adaptation. Bring in large-language-model scenario generation. Add a direct competency-by-language assessment. And run, before any broadening to new user groups and channels, the randomised controlled trial that measures whether behaviour change lasts.

Acknowledgments

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. AP32726003). During the preparation of this work the authors used AI-assisted language and drafting tools; the authors reviewed and edited the content and take full responsibility for the content of the publication.

References

[1] Verizon. (2024). 2024 data breach investigations report. Verizon Business. https://www.verizon.com/business/resources/reports/2024-dbir-data-breach-investigations-report.pdf

[2] Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials*, 15(4), 2091–2121. https://doi.org/10.1109/SURV.2013.032213.00009

[3] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, 10(2), 1–31. https://doi.org/10.1145/1754393.1754396

[4] Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., von Landesberger, T., & Volkamer, M. (2020). An investigation of phishing awareness and education over time: When and how to best remind users. *In Proceedings of the 16th Symposium on Usable Privacy and Security (SOUPS 2020)* (pp. 259–284). USENIX Association.

[5] Canham, M., Posey, C., Strickland, D., & Constantino, M. (2021). Phishing for long tails: Examining organizational repeat clickers and protective stewards. *SAGE Open*, 11(1), Article 2158244021990656. https://doi.org/10.1177/2158244021990656

[6] Soylu, D., Medeni, T. D., Andekina, R., Rakhmetova, R., & Ismailova, R. (2021). Identifying the cybercrime awareness of undergraduate and postgraduate students: Example of Kazakhstan. *In 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)* (pp. 1–6). IEEE. https://doi.org/10.1109/SIST50301.2021.9465995

[7] Government of the Republic of Kazakhstan. (2023). Concept of digital transformation, development of the information and communication technologies industry and cybersecurity for 2023–2029 (Resolution No. 269, March 28, 2023). Astana, Kazakhstan. https://adilet.zan.kz/rus/docs/P2300000269

[8] Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 45(8), 1146–1166. https://doi.org/10.1177/0093650215627483

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

[10] Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

[11] Marshall, N., Sturman, D., & Auton, J. C. (2024). Exploring the evidence for email phishing training: A scoping review. *Computers & Security*, 139, 103695. https://doi.org/10.1016/j.cose.2023.103695

[12] Bitrián, P., Buil, I., Catalán, S., & Merli, D. (2024). Gamification in workforce training: Improving employees' self-efficacy and information security and data protection behaviours. *Journal of Business Research*, 179, 114685. https://doi.org/10.1016/j.jbusres.2024.114685

[13] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

[14] Amjad, K., Ishaq, K., Nawaz, N. A., Rosdi, F., Dogar, A. B., & Khan, F. A. (2025). Unlocking cybersecurity: A game-changing framework for training and awareness — A systematic review. *Human Behavior and Emerging Technologies*, 2025(1), Article 9982666. https://doi.org/10.1155/hbe2/9982666

[15] Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

[16] Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering (EBSE Technical Report EBSE-2007-01). *Keele University and Durham University*. https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[17] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[18] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

[19] Aidynov, T., Goranin, N., Satybaldina, D., & Nurusheva, A. (2024). A systematic literature review of current trends in electronic voting system protection using modern cryptography. *Applied Sciences*, 14(7), 2742. https://doi.org/10.3390/app14072742

[20] van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

[21] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. *In Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS 2007)* (pp. 88–99). ACM. https://doi.org/10.1145/1280680.1280692