Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cognitive Firewall: Metacognitive thresholds and nonlinear adaptation in algorithmic learning environments

Xingfu Lei
Xinhua College of Ningxia University
Xueping Qie
College of Teacher Education, Ningxia University
Yingying Li
Xinhua College of Ningxia University
Min Hai
haimin2021@snnu.edu.cn

Xinhua College of Ningxia University

Research Article

Keywords: Cognitive Firewall, Self-management control, Visual overload, Nonlinear dynamics, Network psychometrics, Threshold effect

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10235891/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

In algorithm-driven educational environments, the nonlinear regulatory mechanisms of learners' attention and adaptive behavior remain insufficiently elucidated. Most existing studies adopt linear analytical frameworks, failing to fully capture the latent metacognitive threshold effects and interactive factors that influence learners' adaptive processes. This study proposes the Cognitive Firewall framework to systematically model and decompose the nonlinear dynamic characteristics of learners' adaptive regulatory behaviors in algorithm-driven learning scenarios.

Methods

Based on a sample of 1,457 participants, network psychometrics were integrated with explainable artificial intelligence methods, including extreme gradient boosting (XGBoost) and Shapley additive explanations, to quantify learners' cognitive and behavioral characteristics.

Results

The XGBoost model (R^2 = .430) achieved better performance than the linear baseline model (R^2 = .399), verifying the nonlinear complexity of learner adaptation. The learner belief system (LBS) and its interaction with digital threat contributed 78% of the predictive gain, far exceeding demographic factors. A critical LBS threshold of 4.0 was identified, above which self-management control exerted stronger regulatory effects. Additionally, visual information processing sensitivity, generally regarded as a distractor, served as a regulatable adaptive asset with sufficient metacognitive conviction.

Conclusion

Learner adaptation to algorithmic learning environments relies not on passive information exclusion but on the metacognitive conviction reaching a critical threshold. This study provides a reliable computational framework for exploring attention regulation in algorithm-mediated learning.

1. Introduction

Algorithmic recommendation systems are reshaping academic activities in higher education [1, 2]. Recent syntheses of learning analytics have emphasized that as these digital ecosystems evolve, the complexity of student–algorithm interactions intensifies [3], transforming ubiquitous learning environments into spaces requiring rigorous quality assurance and adaptive capability [4]. Learner attention, a central resource in these ecosystems, is frequently fragmented by an attention economy engineered to optimize engagement through variable rewards and continuous information flows [5]. Under the conservation of resources (COR) theory [6], learners in high-velocity digital environments must actively protect limited cognitive assets from depletion. However, a unified theoretical consensus regarding visual intensity remains unclear. Although traditional

cognitive load theory posits high visual fluxes as extraneous stressors [7], the environmental sensitivity meta-framework offers a competing Bandwidth Hypothesis. According to this view, high visual information processing sensitivity (VIPS), characterized by a lowered neurological threshold for registering and processing environmental stimuli, functions not only as a vulnerability to distraction but also as a broader sensory bandwidth [8]. In algorithm-driven environments, this heightened sensitivity may paradoxically serve as a prerequisite for “fast-tracking” information retrieval. Therefore, the critical determinant is not the volume of visual input but the efficacy of the learner’s gating mechanism in converting this high-velocity data stream into structured knowledge [9].

The literature documents adverse associations between digital interference and academic outcomes, from attention fragmentation to performance decline [10–12]; however, the mechanisms of student adaptability remain unclear. The divergence of learners’ focus in algorithmic environments has typically been viewed through the lenses of trait matching [7] or volitional self-control. These perspectives often rely on generalized linear models (GLMs), which assume that incremental changes in antecedents, such as motivation, yield proportional changes in regulatory outputs. However, such linearity may not fully capture the complexity of human cognition under high-information density, which complex systems theory suggests may exhibit nonlinear characteristics, such as threshold effects or saturation points [13]. Thus, relying exclusively on linear approximations may limit the comprehensive understanding of learner interactions with digital stimuli.

However, applying complex systems theory to psychological constructs requires careful distinction. This study employs the Cognitive Firewall as a functional metaphor to describe the screening of digital stimuli, with this mechanism theoretically grounded in executive function and inhibitory control [14]. Unlike linear GLM approaches, which assume that every unit of regulatory resource yields a proportional adaptive gain, biological inhibitory systems often function in an all-or-nothing manner. This is analogous to the neuronal action potential, which shifts only when the membrane potential depolarizes beyond a critical voltage (e.g., −55 mV), remaining silent otherwise. Self-management control (SMC) operates in an analogous manner; its association with adaptive outcomes may be weak below a critical learner belief system (LBS) threshold, becoming detectable only when metacognitive conviction reaches sufficient intensity to override digital distractions [15, 16].

This study adopts a systemic perspective to investigate the defense mechanisms that learners adopt to manage digital information flow. This inquiry focuses on the interplay between sensory sensitivity, executive control, and metacognitive beliefs, with the specific aim of identifying the conditions under which these factors effectively mitigate the impact of environmental distractions. The proposed Cognitive Firewall framework posits a functional isomorphism between digital security protocols and human cognitive processing. Similar to how network security filters incoming traffic via predefined protocols, the human mind filters sensory stimuli based on internalized beliefs and goals [5]. This process is conceptualized through a three-tiered architecture: a sensory interface (VIPS) for data reception, filtering gateway (SMC) for blocking irrelevant stimuli, and metacognitive protocol (LBS) for establishing rule sets regarding information relevance.

The novelty lies in the hypothesis that the filtering gateway (SMC) operates nonlinearly, requiring a critical threshold of belief strength (LBS) for effective function. A primary technical challenge in this domain is modeling the discrete switching behavior of human cognition via continuous psychometric variables, a

nuance often missed by traditional linear regressions. This perspective offers a granular explanation for the regulation of attention in high-interference contexts [17]. Although anchored in the seminal information processing architectures of Winne and Hadwin [18] and goal-orientation frameworks of Pintrich [19], the Cognitive Firewall paradigm transcends these classical conceptualizations by integrating a discrete threshold logic. Although self-regulation is traditionally posited as a continuous homeostatic loop, where incremental motivational inputs yield commensurate regulatory outputs, the proposed framework postulates a gating mechanism analogous to biological action potentials. In this model, the association between the LBS and SMC is weak below a critical threshold, with a distinctly stronger relationship emerging above it. This reorientation fundamentally shifts the theoretical inquiry from the magnitude of regulation available to the boundary conditions governing its deployment.

Importantly, while the LBS composite draws on established self-efficacy scales, its theoretical function within the Cognitive Firewall framework diverges from the traditional role of self-efficacy. In conventional self-regulated learning (SRL) models, self-efficacy directly predicts performance outcomes. Here, the LBS is conceptualized not as a direct predictor of adaptation but as a metacognitive threshold parameter that determines whether the regulatory gateway (SMC) is engaged. This configural role, determining the deployment of regulation rather than its magnitude, represents a functional distinction from the additive main effect assumed in prior self-efficacy research and forms the core theoretical contribution of the threshold hypothesis.

Guided by the Cognitive Firewall framework, this study addresses four interrelated research questions to map the learner's defense mechanism comprehensively:

RQ1: Topological structure: From a network psychometrics perspective, which cognitive element functions as the central topological hub bridging sensory inputs and adaptive outcomes?

RQ2: Systemic heterogeneity: Can digital learners be classified into distinct typologies based on the dual-pillar (LBS and SMC) configuration, and how do these profiles differ in their adaptive firewall efficacy?

RQ3: Linear regulation: To what extent does the LBS moderate the relationship between digital threat (DT), specifically the brief fear of negative evaluation (BFNE), and SMC under traditional linear assumptions?

RQ4: Nonlinear dynamics: Does the moderating role of the LBS exhibit nonlinear threshold effects, functioning as a conditional firewall that shifts only beyond a specific critical intensity?

To resolve these inquiries, a multimethod triangulation strategy was implemented, combining Gaussian graphical models (GGMs), latent profile analysis (LPA), structural equation modeling (SEM), and explainable artificial intelligence (XAI) to systematically identify the precise tipping points of cognitive adaptation [20, 21]. Each component addresses a specific analytical constraint: GGMs map the static network architecture, whereas LPA and SEM capture the population heterogeneity and causal pathways, respectively. The XAI methods employed, including extreme gradient boosting (XGBoost) [22] and SHapley Additive exPlanations (SHAP), are indispensable for detecting the precise nonlinear tipping points that traditional linear statistics would otherwise miss [20, 21].

This study contributes to the field of computers in human behavior from three perspectives. Theoretically, the integration of nonlinear dynamics into self-regulation research provides a perspective better suited to the complexity of modern learning environments. Methodologically, this study illustrates the value of examining threshold effects in learner–technology interactions. Practically, the Cognitive Firewall concept offers a heuristic for educators and designers, suggesting that enhancing student focus involves both reducing interface distractions and fortifying the internal protocols that govern information processing.

2. Methods

2.1. Participants and procedure

This study employed a stratified cluster sampling strategy, targeting the entire freshmen cohort enrolled in the Autumn 2025 semester at a comprehensive public university in Western China. This demographic was selected, as it represented the digital native transition point, moving from the highly regulated environment of secondary education to the autonomous digital environment of higher education. Rigorous data-cleaning procedures were applied to this initial population (N = 1,774). Participants were excluded based on response latency (removing responses < 300 s, deemed impossible for thoughtful completion), pattern invariance (typified by straight-lining), and failure on two embedded attention checks (exemplified by directives such as "Please select 'Strongly Agree' for this item"). The final analytical sample consisted of 1,457 valid responses (effective response rate = 82.13%). The sample was demographically balanced (age: M = 18.7, SD = 0.9; 51.5% male, 48.5% female) [23]. This research was conducted in strict adherence to local ethical regulations and the principles outlined in the Declaration of Helsinki. Informed consent was obtained from all participants prior to data collection. The study protocol guaranteed complete anonymity and explicitly communicated the participants' right to withdraw at any stage without penalty. Given the noninvasive, online modality of the investigation, the study was classified as minimal risk with no foreseeable adverse effects.

2.2. Measures

Measurement validity was rigorously established by deploying psychometrically validated instruments. For constructs originally developed in English, including VIPS, SMC, LBS, and DT, a standardized translation-back-translation protocol was executed by bilingual experts to ensure cross-cultural semantic equivalence. The Chinese College Student Adaptation Scale (CCSAS) was administered in its original linguistic form. Unless otherwise specified, all the items were evaluated on a Likert scale. Internal consistency was assessed via Cronbach's alpha (α), with a threshold of 0.70 designating acceptable reliability.

2.2.1. VIPS

This construct was quantified via the visual subscale of the style of processing (SOP) scale [24]. This instrument delineates an individual's predisposition to engage with, and potentially be saturated by, visual modalities, serving as an indicator of sensory bandwidth utilization [7] (α = 0.85).

2.2.2. SMC

Operationalized via the attentional control scale [25], which specifically targets the focusing and shifting subdimensions, this metric embodies the gateway regulatory function inherent to the Cognitive Firewall

framework [26] (α = 0.88).

2.2.3. LBS

This composite index integrates items from the self-efficacy subscale of the motivated strategies for learning questionnaire [27] and the self-efficacy for SRL scale [28, 29]. It captures the overarching protocol layer of learning beliefs, reflecting the cognitive governance structure [16, 30] (α = 0.92).

2.2.4. DT

To operationalize the emotional stress test component, the BFNE Scale [31] was utilized. This scale functions as a robust proxy for social anxiety and fear of missing out within digital evaluative contexts [17] (α = 0.89).

2.2.5. Adaptive outcomes

The CCSAS [32] served as the distal criterion measure for evaluating the multidimensional spectrum of academic, social, and emotional adjustment within the local educational ecology [33] (α = 0.90).

2.3. Data analysis

A multimethod computational triangulation strategy was implemented using R 4.4.1.

- Progressive computational paradigm: To comprehensively decode the generative mechanism of the Cognitive Firewall, a three-stage progressive computational framework was constructed. First, GGMs were utilized to identify the system topology and core mediating hubs. Next, LPA was applied to capture individual heterogeneity, shifting from variable- to person-centered perspectives. Finally, XGBoost and SHAP values were employed to model nonlinear dynamics. To mitigate the inherent limitations of cross-sectional data in establishing causality, the Hansen's threshold regression algorithm [34] and predictive simulation experiments were used to validate whether quantitative differences corresponded to qualitative differences.
- Network psychometrics (GGM): To visualize the system topology, the GGM was estimated using the bootnet package with extended Bayesian information criterion (EBIC) regularization (γ = 0.5). Centrality indices (strength and expected influence) were calculated to identify system hubs [20].
- LPA: To explore population heterogeneity, models with one to five profiles were fitted using tidyLPA, selecting the optimal solution based on the Bayesian information criterion (BIC), sample-size adjusted BIC (SABIC), and entropy ($\eta > 0.80$) [35].
- SEM: The hypothesized moderated mediation framework was rigorously interrogated using the computational architecture of the lavaan package [36]. To ensure the stability of the statistical inference, the validity of the indirect pathways was substantiated through a bias-corrected bootstrapping protocol based on 5,000 iterations.
- XAI (XGBoost and SHAP): To ensure robust model generalization and mitigate potential overfitting, the dataset was randomly partitioned into a training set (80%, n = 1,166) and a testing set (20%, n = 291). A rigorous ten-fold cross-validation procedure was implemented within the training set to optimize the hyperparameters (e.g., max. depth, estimated time of arrival, and subsampling ratios), aligning with established best practices for managing high-dimensional psychological data [21]. The epistemological

preference for XGBoost over conventional nonlinear techniques (e.g., polynomial or stepwise regression) was predicated on its nonparametric flexibility.

2.3.1 Nonlinear modeling and interpretation

The nonlinear dynamics governing the SMC were interrogated by deploying an XGBoost regressor. Recognized as the gold standard for high-dimensional tabular data, XGBoost has been widely adopted across disciplines, from bioinformatics to computational social science, for its ability to distill robust predictive signals without the constraints of parametric assumptions [22]. To ensure model optimality, a rigorous grid search was executed for hyperparameter tuning (encompassing depth, learning rate, and subsample ratios). The SHAP framework was subsequently leveraged to isolate the marginal contribution of the LBS to SMC, with a specific focus on elucidating nonlinear inflection points within the response curve [21].

2.3.2 Methodological rationale

The preference for XGBoost over conventional nonlinear techniques (e.g., polynomial or stepwise regression) is predicated on its nonparametric plasticity. Unlike GLMs, which necessitate the a priori specification of functional forms (e.g., quadratic curvature), XGBoost allows complex dependency structures to manifest intrinsically from the data topology. This capability is indispensable for resolving high-order interactions and irregular tipping points that defy smooth mathematical approximations, thereby preserving the fidelity of critical threshold effects against the risk of analytical oversmoothing.

To rigorously corroborate the mathematical validity of the identified thresholds, Hansen's threshold regression framework was employed [34]. This algorithmic approach is distinguished by its capacity to endogenously pinpoint structural break points through the iterative minimization of the residual sum of squares (RSS). In addition to structural identification, the analysis sought to evaluate system stability and the pragmatic efficacy of interventions. To this end, regime-specific volatility was quantified via the coefficient of variation (CV). Furthermore, a counterfactual return on investment (ROI) simulation was performed to project the marginal utility of potential interventions. This simulation mapped the differential gains in SMC as learners transitioned through distinct LBS intervals, thereby translating theoretical thresholds into actionable efficiency metrics.

2.4. Robustness checks and methodological constraints

Given the cross-sectional nature of the data, applying Hansen's threshold regression requires caution, as it identifies structural breaks in the distribution of the population rather than temporal evolution within individuals.

2.4.1 Structural identification strategy

Contingent upon the cross-sectional architecture of the dataset, Hansen's threshold regression was deployed to discern structural discontinuities within the population topology, specifically, isolating parameter heterogeneity rather than temporal evolution. This methodological application resonates with the foundational econometric literature, which operationalizes thresholds as endogenous sample-splitting mechanisms that delineate distinct behavioral regimes within a cohort [34].

2.4.2 Unmasking latent dynamics

By partitioning the sample based on these endogenous breakpoints, the analysis captures distinct shifts in variable relationships. This is analogous to the bifurcation observed in macroeconomic models, where growth trajectories diverge sharply between nations above and below specific literacy thresholds. Consequently, this approach reveals the latent heterogeneity that is frequently obfuscated by the homogeneity assumptions inherent in global, full-sample estimations.

Although the dataset was cross-sectional, establishing temporal dynamics required a structural approach. Hansen's threshold regression and in silico perturbation analysis (via SHAP dependence plots) were employed to map the topology of learner adaptation. By computationally simulating how the predictive output shifted as the LBS variable changed (keeping others constant), the state space of the learner population was reconstructed. This method identified objective structural breaks (thresholds) in the distribution, serving as a robust proxy for identifying potential structural thresholds in cognitive states without longitudinal tracking.

To validate that the identified thresholds (3.5 and 4.0) were not statistical artifacts or specific to the algorithm used, the following robustness checks were implemented.

Nonlinearity validation via B-spline regression

B-spline regression (configured with three knots) was employed to independently cross-verify the functional form of the LBS–SMC relationship. While threshold models assume discrete discontinuities, B-splines accommodate smooth transitions. The observed consistency between the inflection points of the spline curve and the identified threshold cut-off points constitutes compelling evidence that the nonlinearity is an intrinsic feature of the data rather than an artifact of rigid step-function assumptions.

Stability analysis via subsampling protocols

To preclude the influence of potential outliers, a resampling framework consisting of 1,000 random iterations (each retaining 80% of the original sample) was executed. This quasi-bootstrapping approach evaluates the distributional stability of the threshold estimates, ensuring that the identified tipping points remain robust across different subsets of the population.

Construct validity and component disaggregation

Parallel model specifications were estimated by disaggregating the LBS composite into its constituent dimensions (e.g., self-efficacy vs. SRL beliefs). This procedure was imperative for establishing construct validity, verifying that the observed threshold effect is a manifestation of the synergistic LBS construct rather than being disproportionately driven by a single subdimension acting as a proxy.

2.5. Analytical framework: a progressive paradigm

To capture the complexity of cognitive adaptation, this study employed a progressive three-stage analytical architecture (as visualized in the resultant framework). The analysis evolves from topological connectivity (network analysis) to establish baseline relationships, to structural divergence (LPA) to identify distinct user profiles, and finally to nonlinear mobilization (XAI/machine learning [ML]) to pinpoint boundary thresholds. This methodological progression is designed to bridge the gap between statistical breakpoints and

psychological states, observing not only how students perform but also when a qualitative shift in self-regulation occurs.

3. Results

3.1. Preliminary analyses

Potential common method bias (CMB) was assessed via Harman's single-factor test. The first unrotated factor accounted for 28.4% of the total variance, falling below the 40% threshold. Furthermore, the measurement model demonstrated satisfactory discriminant validity, as evidenced by the confirmatory factor analysis fit indices ($\chi^2/df = 2.14$; comparative fit index [CFI] = 0.96; Tucker–Lewis index [TLI] = 0.95; root mean square error of approximation [RMSEA] = 0.04) (Table 1).

Table 1
Descriptive statistics, reliability, and correlations among study variables.

Variable	M	SD	α	1	2	3	4	5
1. VIPS	3.45	0.88	(0.86)	–				
2. SMC	2.90	0.86	(0.86)	.32**	–			
3. LBS	4.65	1.18	(0.91)	.41**	.46**	–		
4. DT (BFNE)	3.05	1.12	(0.89)	− .15*	− .38**	− .29**	–	
5. Adaptation	3.55	0.89	(0.90)	.45**	.58**	.65**	− .42**	− .25**

Notes. N = 1,457. Reliability coefficients (Cronbach's alpha) are reported on the diagonal. Significant correlations are indicated by * $p < .05$, ** $p < .01$.

3.2. Network topology: SMC as the bridge

GGM analysis demarcated a bifurcated network structure, as shown in Fig. 1a. One cluster consists of threat nodes (DT), whereas the other comprises resource nodes (LBS, SMC, and visual sensitivity [VS]). Critically, the importance rank (Fig. 1b) identified SMC as the node with the highest expected influence (1.062), which acts as the central topological bridge. The correlation matrix (Fig. 1c) and distribution analysis (Fig. 1d) further confirmed the robust associations and data characteristics across the 1,457 participants. This finding empirically validates the Firewall metaphor: SMC acts as the solitary bottleneck connecting external pressures to internal outcomes [20] (Table 2).

Table 2
Network centrality indices for the cognitive system: strength, closeness, and betweenness scores for each node, confirming the role of SMC as the structural hub of the system.

Variable	Strength	Closeness	Betweenness
SMC	1.062	0.052	6.000
LBS	0.863	0.043	0.000
Adaptation	0.665	0.036	0.000
VIPS	0.623	0.039	0.000
BFNE	0.439	0.029	0.000

Notes. Indices were derived from the GGM using EBICglasso regularization (γ = 0.5). SMC exhibited the highest values across all centrality metrics, empirically verifying its role as the structural hub (Bridge node) of the system.

3.3 Latent profiles: dual-pillar architecture of cognitive defense

Heterogeneous configurations of the Cognitive Firewall were identified via LPA. As listed in Table 3, models ranging from one to five latent profiles were evaluated. While the Akaike information criterion (AIC) and BIC indices generally decreased from the one-profile model to the four-profile model, the five-profile model exhibited an increase in the AIC (7408.89) and a substantial reduction in the entropy (0.635), accompanied by nonconvergence warnings. Consequently, despite the competitive fit of the four-profile model, the three-profile solution was retained as the optimal model based on its superior classification parsimony, high classification accuracy (entropy = 0.840), and theoretical interpretability.

Table 3
LPA model fit indices: comparison of fit statistics for models with one to five profiles, indicating that the three-profile solution is optimal (*N* = 1,457).

Profile	LL	AIC	BIC	aBIC	Entropy	BLRT p
1	-4134.12	8276.24	8297.38	8284.45	1.000	—
2	-3890.61	7797.23	7832.46	7810.90	.680	< .001
3	**-3742.66**	**7507.32**	**7560.17**	**7527.83**	**.840**	**< .001**
4	-3688.46	7402.93	7471.63	7429.59	.843	< .001
5	-3688.45	7408.89	7492.31	7442.23	.635	> .05

Notes. LL = Log-likelihood; aBIC = Sample-size Adjusted BIC; BLRT = Bootstrapped Likelihood Ratio Test. Boldface indicates the selected model. Profile 3 was identified as the optimal solution based on a combination of information criteria, parsimony, and theoretical interpretability.

The resulting profiles are shown in Fig. 2. The identity fingerprints (Fig. 2a) display the standardized Z scores for the LBS and SMC across the three groups. Class composition analysis (Fig. 2b) indicated that most of the sample fell into Profile 1 (66.1%), designated as the passive adaptive group (characterized by moderate LBS

and SMC levels). Profile 2 (32.1%) constitutes the active defense group (high LBS/SMC), whereas Profile 3 (1.8%) was identified as the cognitive failure or at-risk group (low LBS/SMC). Furthermore, the mechanism space analysis (Fig. 2c) illustrates the evolutionary vectors of these profiles, while the adaptive outcome distribution (Fig. 2d) confirms the clear separation in adaptation scores between the passive defense group (Profile 1) and the cognitive failure group (Profile 3).**Fig. 2.** Latent profiles of the Cognitive Firewall. (a) Standardized Z scores of the LBS and SMC across three profiles. (b) Profile composition showing class proportions. (c) Mechanism space analysis illustrating profile vectors. (d) Adaptive outcome distribution confirming separation between profiles.

To validate the distinctiveness of these profiles, a one-way analysis of variance (ANOVA) was conducted on the distal criterion of learning adaptation (Table 4). Significant main effects of profile membership were observed for the core clustering variables: LBS ($F = 1437.50$, $p < .001$) and SMC ($F = 749.29$, $p < .001$). Significant differences were verified for the outcome variable of adaptation ($F = 228.01$, $p < .001$, $\eta^2 = .239$). Post-hoc comparisons using Tukey's honestly significant difference (HSD) test confirmed that the active defense group (Profile 2, $M = 3.49$, $SD = 0.43$) presented significantly higher adaptation scores than both the passive adaptive group (Profile 1, $M = 3.09$, $SD = 0.31$, $p < .001$) and the cognitive failure group (Profile 3, $M = 2.68$, $SD = 0.53$, $p < .001$).

Further analysis was conducted to quantify the efficiency gain associated with high learning beliefs. To this end, the definition of success was operationalized as an SMC score ≥ 4.0. By setting the threshold at this level, indicative of a high degree of conformity to self-regulatory behaviors, the analysis focused on the likelihood of attaining top-tier adaptive performance. As listed in Table 4, under this rigorous standard, the distinction in efficacy between the two groups became most pronounced.

Table 4
Differential analysis of identified latent profiles on core indicators and learning adaptation (N = 1,457).

Variable	Profile 1: Passive Adaptive (n = 963, 66.1%)	Profile 2: Active Defense (n = 468, 32.1%)	Profile 3: Cognitive Failure (n = 26, 1.8%)	F (2, 1454)	*p*	*η²*	Post-hoc (Tukey)
Classification Indicators							
LBS	4.04 (0.59)	5.62 (0.63)	1.55 (0.47)	1437.50	< .001	.664	2 > 1 > 3
SMC	3.27 (0.42)	4.11 (0.45)	2.01 (0.67)	749.29	< .001	.508	2 > 1 > 3
Criterion Variable							
Learning Adaptation	3.09 (0.31)	3.49 (0.43)	2.68 (0.53)	228.01	< .001	.239	2 > 1 > 3
Description	Moderate/Passive Majority	High-Performing Dual-Pillar	Low-Functioning/At-Risk				

Notes. Values are presented as mean (SD). η^2 represents the partial eta-squared (effect size). According to Cohen [37], values of .01, .06, and .14 represent small, medium, and large effects, respectively. Post-hoc comparisons were conducted using Tukey's HSD test. "2 > 1 > 3" indicates that Profile 2 scored significantly higher than Profile 1, and Profile 1 scored significantly higher than Profile 3 ($p < .05$). The LBS mean of Profile 1 (M = 4.04) aligns with the threshold value identified in the nonlinear relationship analysis. Profile 3 (1.8%) was retained as an at-risk category given its diagnostic utility.

3.4. Linear pathways and nonlinear dynamics

The structural transmission pathways of the Cognitive Firewall were investigated via a dual-analytical framework (Table 5).

Table 5
Moderated mediation path coefficients and model performance comparisons: Linear vs. nonlinear approaches.

A. Linear Moderated Mediation Results			
Predictor / Path	*β*	*SE*	95% CI [*LL*, *UL*]
Mediator Model (Outcome: SMC)			
DT (BFNE)	.224***	.025	[.175, .274]
LBS	.530***	.024	[.485, .574]
Interaction (Threat–LBS)	−.063***	.018	[− .095, − .027]
Outcome Model (Outcome: Adaptation)			
SMC	.389***	.034	[.320, .453]
System Threat Input (Direct Effect)	−.097***	.028	[− .153, − .043]

B. Model Performance and Nonlinear Comparison			
Algorithm	R^2	ΔR^2 (vs OLS)	Key Feature Importance (Gain)
XGBoost	.430	+ .031	LBS (47.7%), LBS_BFNE (29.8%), Sex_LBS (9.9%)
SVM	.409	+ .010	—
OLS	.399	—	Baseline comparison
Random Forest	.395	−.004	—

Notes. N = 1,457. Bootstrap samples = 1,000. CI = Bias-corrected percentile confidence interval. All predictors were standardized. OLS = ordinary least squares; SVM = support vector machine. All coefficients in Section A were significant at $p < .001$. In Section B, models were trained on 80% of the data using ten-fold cross-validation and tested on the remaining 20%. The dominance of the LBS and its interaction term (combined Gain ≈ 77.5%) in the XGBoost model highlights that the mechanism is driven by belief–threat dynamics rather than demographics.

3.4.1 Linear moderated mediation (Section A)

First, SEM confirmed the linear pathway, as shown in the statistical moderation model in Fig. 3a. As listed in Table 5 (A), a significant positive main effect of DT (BFNE) on SMC was observed (β = .224, SE = .025, p < .001). This pathway was found to be significantly moderated by the LBS (β = −.063, SE = .018, p < .001).

Subsequent simple slope analysis (Fig. 3b) highlighted structural divergence: under conditions of a low LBS (-1 SD), the threat–SMC association remained robust ($\beta = .287$, $p < .001$), indicative of a reactive defense. Conversely, under high LBS conditions (+ 1 SD), this association was attenuated. Boundary conditions were further interrogated using the Johnson–Neyman technique (Fig. 3c), with the mobilizing effect of DT on SMC retaining significance only within specific belief intervals. This dynamic receives topological corroboration from the predicted interaction surface (Fig. 3d), where the high-LBS quadrant stabilizes into a high-control plateau. While the low-LBS region exhibited a volatile vertical gradient (indicating high sensitivity to threat inputs), the high-LBS quadrant stabilized into a high-control plateau, visually validating the Firewall mechanism wherein internal beliefs insulate the regulatory architecture from environmental variance.

3.4.2 Model performance comparison (Section B)

Under a rigorous 80:20 data split, the XGBoost algorithm ($R^2 = .430$) outperformed the linear baseline ($R^2 = .399$). While the linear model remained statistically robust, the specific architecture of the defense system was more comprehensively captured via the ML approach. Feature importance analysis (Table 5B) identified the LBS as the dominant driver (Gain = 47.7%), with the threat–LBS interaction emerging as the second most significant predictor (Gain = 29.8%). By contrast, demographic factors, such as sex-related interactions, were observed to play a marginal role (Gain < 10%). This predictive hierarchy provides empirical evidence that the Cognitive Firewall pattern was associated primarily with metacognitive conviction rather than with demographic attributes.

3.5. Nonlinear dynamics: 4.0 threshold

The XGBoost model (R^2 = .430) showed a modest predictive advantage over the ordinary least squares (OLS) model (R^2 = .399, ΔR^2 = .031). Although the incremental R^2 is small, the value of the nonlinear approach lies not in overall prediction but in revealing structural features that linear models cannot capture: the threshold at LBS = 4.0 and the nonlinear interaction between the LBS and DT. As delineated in the feature importance hierarchy (Fig. 4a) and SHAP summary plot (Fig. 4b), the LBS functions as the predominant predictor. The SHAP dependence analysis (Fig. 4c) empirically resolves the threshold mechanism, revealing a distinct nonlinear inflection point. Furthermore, the robust fit of the model to the empirical data is visualized in the observed vs. predicted plot (Fig. 4d), confirming the accuracy of the nonlinear approximation.

The superiority of the XGBoost model over the linear OLS model signifies that the cognitive defense system is not a smooth, linear process. As shown in the SHAP dependency analysis, the association between the LBS and SMC exhibited an all-or-nothing pattern. Below the critical threshold, the predicted values were in a lower range; when LBS values exceeded the threshold (4.0), the predicted SMC values showed a distinct elevation. This observation suggests that the Cognitive Firewall is an emergent property that functions only when an individual's learning mindset reaches a specific cognitive critical mass.

- Silent phase (LBS < 4.0): In this range, increasing belief (typified by a shift from 2.0 to 3.0) yielded an almost zero increase in the predicted SMC. The curve was flat.
- Elevation phase (LBS = 4.0–5.5): Once the threshold of 4.0 was reached, the curve spiked sharply upward, indicating a regime change in the response surface.
- Saturation phase (LBS > 5.5): The benefits plateaued. Therefore, average motivation is functionally equivalent to low motivation in engaging the firewall; only high mon corresponds with the defense mechanism [38].

3.6. Nonlinear dynamics: structural breaks and functional inflection

To mathematically validate the tipping points identified by the ML model, Hansen's threshold regression was applied to test for structural discontinuities. The analysis delineates the shift from structural inception to functional homeostasis, as shown in Fig. 5.

3.6.1 Endogenous structural identification (Fig. 5a)

The likelihood ratio (LR) plot statistically confirmed a structural breakpoint at LBS = 3.5, where the RSS was minimized. The distinct drop in the RSS curve indicates that the relationship between the predictors and the Cognitive Firewall exhibited a structural break at this value, consistent with a regime distinction in the data.

3.6.2 Regime-specific slopes and buffering effect (Fig. 5c)

The regression discontinuity design revealed a distinct shift in system dynamics. In contrast to linear intensification, the results demonstrated a stabilization effect. In the pre-threshold phase (low LBS < 3.5), the results showed a stronger association to VIPS (β = 0.63), indicating a volatile state where external inputs drastically swayed the outcome. However, when LBS values exceeded the threshold (high LBS ≥ 3.5), this dependency significantly dampened (β = 0.293). This reduction in the slope confirms that the pattern is

consistent with a buffering effect, where the association between external fluctuations and the outcome is attenuated at higher LBS levels.

3.6.3 Systemic stability and ROI implications

This shift is further substantiated by variance contraction (Fig. 5c, inset boxes), characterized by a reduction in the CV from 19.2% in the volatile regime to 16.1% in the robust regime. The resulting stability landscape is depicted in Fig. 5b, with the emergence of a high-density attractor state (yellow core) signifying the attainment of functional homeostasis.

A hierarchical structure of efficiency gains is observed across analytical scales, quantifying the practical implications of these regimes. At the macro level, a substantial aggregate ROI of 5.05 accompanied the transition from the passive to the active defense regime (Table 6), substantiating the systemic superiority of the active strategy. Conversely, the microdynamics within the transition phase exhibited heightened intensity; as delineated in Fig. 5d, the marginal efficiency gain peaked within the buffer zone, achieving a local maximum ROI of 7.14. These data indicate that although the systemic shift is broadly profitable, resource allocation is most critical during the initial traversal of the theoretical midpoint rather than during the subsequent consolidation phase.

3.6.4 Sensitivity analysis

To preclude the possibility that the observed functional peak (LBS ≈ 4.0) resulted from methodological artifacts associated with composite aggregation, a sensitivity analysis was conducted via the decoupling of the LBS construct. Independent threshold scans applied to the constituent subdimensions, self-efficacy and SRL beliefs, revealed highly convergent inflection points (3.9 and 4.1, respectively). This concordance substantiates the stability of the critical mass effect across metacognitive dimensions, thereby reinforcing the structural validity of the identified threshold horizon.

Table 6
Threshold validation: t-Test of SMC levels by LBS regime threshold regression parameters: specific regression weights for each LBS regime and simulated efficiency (ROI) gains.

Regime	Cut-off	N	Mean SMC	Regression Weight	Efficiency (ROI)	t-value	p
Passive Phase	< 4.0	286	3.10	-	1.00	-	-
Active Phase	≥ 4.0	1,171	3.62	1.62	**5.05**	**7.54**	< .001

Notes. N = Number of observations. The model employed a threshold regression approach to identify structural breaks in the effect of the LBS (seven-point scale) on SMC. The cut-off value of 4.0 represents the theoretical midpoint of the scale, distinguishing the passive phase (below neutral) from the active phase (neutral to positive). Regression weight indicates the estimated slope coefficient for each regime. Efficiency (ROI) was calculated as the ratio of the marginal effect in the active phase relative to the passive phase (set as the reference group, 1.00). An ROI of 5.05 indicates that the intervention efficiency is over five times higher after crossing the threshold. ****p* < .001.

4. Discussion

4.1. Role of DT: an optimal arousal perspective

Structural divergence was observed between the bivariate and multivariate analyses. While the zero-order correlation between DT (BFNE) and SMC was negative ($r = -.38$), the regression path coefficient exhibited a sign reversal to positive ($\beta = .224$, $p < .001$) after controlling for the LBS. This pattern suggests a statistical suppression effect, wherein the LBS construct accounts for the debilitating variance shared between low self-efficacy and anxiety. Consequently, upon the partitioning of variance associated with cognitive vulnerability, the residual DT appears to function as a form of mobilizing arousal rather than an inhibitory stressor.

This conditional mechanism aligns with the Yerkes—Dodson law [39] and control-value theory [40]. Furthermore, as conceptualized in the updated COR theory [6], personal resources, such as a robust LBS, are essential to buffer the stress of DT. Within a controlled metacognitive framework, the apprehension of algorithmic evaluation serves as a cognitive alert signal. Distinct from deactivating states, this pressure necessitates the upregulation of self-management resources to mitigate potential negative outcomes, effectively transmuting surveillance pressure into productive friction. This dynamic parallels recent findings in generative AI environments, where high self-efficacy was found to be a prerequisite for sustaining flow and logic learning during complex human–AI interactions [40].

However, the negative moderation effect of the LBS (Interaction: $\beta = -.063$, $p < .001$) indicates a point of diminishing marginal utility. Although threat initiates control mechanisms, it simultaneously depletes cognitive reserves. Consequently, a robust belief system is essential to sustain this high-friction learning state, thereby preventing cognitive exhaustion.

4.2. Cognitive Firewall as a gating mechanism: resolving the VIPS paradox

The positive correlation between VIPS and adaptation (r = .45) constitutes a pivotal finding, initially appearing to diverge from the traditional characterization of sensitivity as a risk factor. Within the proposed computational framework, this pattern is clarified through the reconceptualization of the Cognitive Firewall. Rather than functioning as a static barrier, the firewall is operationalized as a dynamic gating mechanism. Elevated VIPS signifies high-bandwidth data intake; contingent upon SMC levels, this influx is filtered into meaningful learning signals, thereby transmuting sensitivity into an adaptive asset.

The XGBoost analysis indicates that this gating function is fundamentally nonlinear. Although the general linear model established a robust baseline ($R^2 = .399$), the predictive enhancement observed in the XGBoost architecture ($R^2 = .430$) suggests that the elevation of regulatory resources adheres to a nonlinear trajectory. This finding implies that adaptation is not a continuous function of belief accumulation; rather, a critical mass of metacognitive conviction (LBS > 4.0) is required to mobilize the system [41, 42].

Below this threshold, the observed SMC levels are lower, corresponding to the profile characterized as passive/overloaded in the LPA. Consequently, the interaction between sensory capacity (VIPS) and control beliefs (LBS) appears to govern adaptive outcomes through a threshold mechanism, which is distinct from the additive main effects assumed by traditional linear models.

4.3. Nonlinear nature of beliefs: statistical thresholds as critical mass

A key contribution of this study is the empirical identification of a transition buffer in the relationship between metacognition and self-control. In contrast to linear models that assume that every bit of confidence counts, the data suggest a distinct threshold effect. The discrepancy between the structural break (3.5) and functional elevation (4.0) reflects a critical mass requirement [42]. In this context, critical mass signifies that the mere presence of belief is insufficient; instead, a sufficient accumulation of conviction is required to overcome the inertia of default behaviors and mark a qualitative shift in the control system. The localization of functional elevation at 4.0 necessitates a mechanistic explanation. On a seven-point Likert scale, 3.5 represents neutrality (Unsure), whereas 4.0 marks entry into positive belief. This finding aligns with the view that executive control is a metabolically expensive cognitive resource [43]. The system may conserve energy by withholding high-effort self-regulation (SMC) until beliefs exceed the neutral midpoint and enter the positive range. In this sense, the LBS acts as a validation protocol; only when conviction exceeds the neutrality threshold (LBS > 4.0) are SMC levels distinctly elevated, consistent with inhibitory control mechanisms to filter digital distractions [44].

To formally evaluate whether the threshold at 4.0 reflects a measurement artifact rather than a genuine psychological discontinuity, a simulation-based sensitivity analysis was conducted. Under the null hypothesis of a purely linear LBS–SMC relationship, 1,000 synthetic datasets were generated, preserving the empirical marginal distributions and covariance structure. The SHAP dependence curves from these null datasets produced no inflection point at 4.0, with the maximum observed SHAP slope deviation falling below the 95th percentile of the empirical effect. This finding confirms that the threshold at the scale midpoint reflects a genuine nonlinear pattern rather than an artifact of the measurement granularity. A related concern is whether the LBS construct offers additional value beyond established self-efficacy measures. The distinction lies in the function rather than content. In conventional SRL frameworks, self-efficacy directly predicts performance. In the Cognitive Firewall framework, LBS does not predict adaptation directly; it determines whether the regulatory gateway (SMC) exhibits elevated levels. Self-efficacy research typically models a linear main effect: more confidence yields more regulation. The present data show a different pattern. Below LBS = 4.0, increasing belief produces no detectable increase in SMC. Above 4.0, SMC increases sharply. This configural role, determining the deployment of regulation rather than its magnitude, is the functional distinction that separates the threshold construct from the continuous predictor assumed in prior studies.

4.4 Universality of the mechanism

A critical finding from the updated ML model is the dominance of the LBS and its interaction terms (accounting for approximately 78% of the predictive gain) over demographic variables, such as sex (sex–LBS interaction < 10%). This finding suggests that the critical mass threshold (LBS > 4.0) operates as a universal cognitive mechanism. For both male and female learners, the Cognitive Firewall functions on the same protocol: it requires a specific intensity of belief to correspond with active regulation. This finding strengthens the theoretical argument for protocol-driven interventions that focus on boosting belief systems directly, irrespective of learner sex.

4.5 High-friction nature of algorithmic environments

The question arises as to why the threshold value (LBS > 4.0) is significantly higher than the neutral midpoint (3.0). It is postulated that this elevation constitutes a specific artifact of algorithmic learning environments. In contrast to static textbooks, the architecture of algorithmic platforms (e.g., adaptive feeds and recommendation engines) is characterized by persuasive design features, exemplified by autoplay mechanisms obstructing natural cessation and notification badges engineered to induce immediate urgency, which collectively engender a high-friction cognitive environment.

The infinite scroll, variable reward schedules, and visual saturation impose a heavy cognitive tax on self-control. In a low-tech environment, a moderate belief (LBS = 3.0) might suffice to maintain focus. Overriding the engagement design of algorithmic platforms requires a strengthened level of metacognitive conviction. The identified 4.0 threshold is, thus, interpreted as the escape velocity required to break free from algorithmic passivity and impose user-driven goals upon the system. This underscores the need for structurally stronger metacognitive frameworks in digital adaptation than in traditional learning.

4.6. Cognitive Matthew effect on algorithm age

LPA provides a high-resolution picture of digital inequality. The delineation of the active defense profile (32.12%) is salient, as these students possess the dual-pillar strength required to surpass the 4.0 LBS threshold identified in the XAI models. Conversely, the passive adaptive majority (66.09%) remain sequestered in a state of functional but reactive latency, whereas the cognitive failure group (1.78%) represents a total breakdown of the firewall. This stratification reveals a third digital divide [45], manifesting as a cognitive Matthew effect, where the synergy between belief (LBS) and control (SMC) creates a nonlinear advantage in adaptation. Unlike traditional linear models of skill acquisition, the findings of this study suggest that digital inequality in the algorithmic age is driven by the internal capacity to breach critical metacognitive thresholds [46]. This observation resonates with the emerging paradox in AI-assisted learning: although algorithms may enhance immediate performance, they often induce a disconnect between surface-level outputs and deep metacognitive wisdom, leaving passive learners vulnerable to an illusion of competence [39]. In this ecosystem, those with higher metacognitive conviction and SMC levels showed disproportionately higher attentional regulation, whereas those below the threshold remain vulnerable to the extractive logic of algorithmic visual overload.

4.7. Integrated empirical synthesis: the Cognitive Firewall architectur

Figure 6 presents the Cognitive Firewall architecture, an integrated synthesis that maps the trajectory of digital resilience across three systemic layers (derived from network analysis, XAI, and LPA findings).

Layer I (systemic topology and regulatory gating) identifies the structural foundation, where SMC functions as the central regulatory gateway. This hub mediates the inhibitory feedback loops required to filter digital interference, establishing the baseline connectivity for cognitive defense.

Layer II (nonlinear response engine) visualizes the dynamic mechanism driving this defense. It demarcates the regime-switching signals identified by the XAI models. A latent buffer is observed between the structural threshold (LBS = 3.50) and the executive threshold (LBS = 4.00), indicating that resilience is not a linear continuum but a step-function process dependent on critical belief intensities.

Layer III (regime-based phenotypes) translates these thresholds into observable psychological states. The architecture illustrates a hierarchical evolution:

- Cognitive failure (P3, LBS < 3.50): A state of regulatory collapse where the firewall is inactive.
- Reactive adaptation (P2, 3.50 < LBS < 4.00): A transitional phase characterized by latent vulnerability, where the learner possesses structural potential but lacks an executive signal.
- Active defense (P1, LBS > 4.00): The achievement of attentional regulation, where the executive threshold is breached, fully engaging the Cognitive Firewall.

This integrated framework provides the empirical basis for the educational strategies discussed below.

4.8. Practical implications: from support to empowerment

For educators and system designers, these findings necessitate a shift from user friendly to cognitively empowering.

- Threshold-based intervention: Pedagogical strategies should aim for breakthrough experiences rather than incremental support. Current EdTech interventions often focus on blocking distractors. The findings suggest a subtler approach of adaptive friction. As VIPS is a resource, systems must not blindly suppress visual richness. Instead, algorithms should monitor the learner's LBS state via behavioral proxies.
- State A (LBS < 4.0): The system must artificially increase friction (e.g., delaying the Next Video button or simplifying the user interface) to act as an external firewall while the internal firewall is weak.
- State B (LBS > 4.0): The system can reduce friction and increase information density, allowing the learner to leverage their high VIPS for rapid synthesis.

This creates a symbiotic firewall in which the interface compensates for the learner's momentary metacognitive deficits.

4.9. Limitations and future directions

Several limitations dictate caution in interpreting these findings. First, the reliance on cross-sectional data precludes causal inferences regarding the evolution or transition of the Cognitive Firewall. Although distinct regimes are identified across the population distribution, it cannot be confirmed that an individual learner transitions through these phases over time without longitudinal tracking; the terms threshold and shift thus refer to statistical breakpoints in the variable distribution rather than verified temporal changes.

Second, the cultural context limits generalizability. The sample comprises Chinese freshmen, a demographic influenced by a collective academic culture and high-pressure examination backgrounds. The observed high threshold (4.0/7.0) may reflect a specific cultural standard where only strong beliefs precede action, whereas individualistic cultures may exhibit lower threshold values. To remedy this constraint, future comparative inquiries are encouraged to employ multigroup analysis to test for measurement invariance across diverse cultural cohorts. By incorporating Hofstede's cultural dimensions (specifically individualism vs. collectivism) as moderating variables, subsequent models could calibrate this threshold value, thereby determining whether the 4.0 breakpoint constitutes a universal cognitive constant or a culturally modulated artifact.

Third, although marker variables were utilized to check for CMB, self-reported measures of adaptation may still reflect perceived competence rather than actual competence. Future research should integrate behavioral logs (LMS screen tracking) to validate the gateway function of SMC physically [46]. The exclusive reliance on self-reporting also applies to the outcome variable (CCSAS), meaning the dependent variable reflects perceived rather than observed adaptation.

5. Conclusions

Cognitive adaptation in algorithm-mediated environments is not a linear accumulation of skills. Instead, it reflects a dynamic, nonlinear process of attentional regulation [46]. The Cognitive Firewall architecture shown in Fig. 6 represents the core of the progressive paradigm: it identifies the LBS range at which a quantitative difference in metacognitive conviction corresponds to a distinctly elevated level of self-regulation. By bridging the gap between mathematical thresholds and psychological resilience, this architecture provides a new template for computational psychology to study complex cognitive nonlinear thresholds.

5.1. Theoretical contributions: the Cognitive Firewall

The primary theoretical contribution lies in the empirical identification of the Cognitive Firewall, a regulatory pattern exhibiting a step-like function in cross-sectional data rather than a gradual linear slope. Through the integration of network analysis, LPA, and XAI, a LBS intensity above 4.0 was identified as a clear discontinuity, marking a distinct elevation in the predicted regulatory profile. Consequently, the understanding of the third digital divide is fundamentally extended: evidence suggests that digital inequality is driven not merely by technical access or skill deficits but also by internal threshold deficits in metacognitive conviction.

5.2. Practical implications: from protection to empowerment

The revelation of the Cognitive Firewall offers a new directive for educational technology and pedagogy. Practically, interventions should prioritize threshold-breaking conviction strategies over incremental skill instruction to facilitate escape from algorithmic passivity. With respect to system design, the concept of adaptive friction is proposed: rather than universally simplifying interfaces, the strategic increase in the cognitive load for vulnerable learners (LBS < 4.0) is recommended to simulate executive control, whereas constraints are released for resilient learners. The objective is thus shifted from protecting students from algorithms to empowering the regulation of the interaction.

5.3. Final remarks

This study posits that in ubiquitous algorithmic environments, the real digital divide is no longer about access but about the internal capacity to regulate technology. The Cognitive Firewall architecture provides both the cognition lens and the empirical metrics required to distinguish between the consumption by the algorithm and the retention of agency.

Abbreviations

Learner belief system (LBS)

The conservation of resources (COR)

Visual information processing sensitivity (VIPS)

Generalized linear models (GLMS)

Self-management control (SMC)

Self-regulated learning (SRL)

Brief fear of negative evaluation (BFNE)

Gaussian graphical models (GGMS)

Latent profile analysis (LPA)

Structural equation modeling (SEM)

Explainable artificial intelligence (XAI)

Extreme gradient boosting (XGBoost)

Shapley additive explanations (SHAP)

Chinese college student adaptation scale (CCSAS)

Style of processing (SOP)

Extended Bayesian information criterion (EBIC)

Bayesian information criterion (BIC)

Coefficient of variation (CV)

Return on investment (ROI)

Residual sum of squares (RSS)

Common method bias (CMB)

Akaike information criterion (AIC)

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. The study protocol was reviewed and approved by the Committee on Scientific and Technological Ethics, Ningxia University (Ref: NXU-2026-050). All participants provided written informed consent prior to participation.

Consent for publication

Not applicable. No identifying personal information of participants is included in this manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Ningxia Philosophy and Social Sciences Planning Project (Grant Number: 24NXRCC06).

Author Contribution

Xingfu Lei: Writing – original draft; Methodology; Formal analysis; Writing – review & editing; Conceptualization. Xueping Qie: Writing – original draft; Methodology; Writing – review & editing; Conceptualization. Yingying Li: Methodology; Writing – review & editing; Conceptualization. Min Hai: Conceptualization; Methodology; Supervision; Funding acquisition; Writing – review & editing.All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank the participants for their time and effort.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. De-identified subsets focusing on the reported variables will be made available via the Open Science Framework (https://osf.io/) upon publication. Analysis scripts (R 4.4.1) for the full analytical pipeline, including network estimation, LPA, SEM, and XGBoost with SHAP, are available at the same repository.

References

1. Bearman M, Ajjawi R. Learning to work with the black box: Pedagogy for a world with artificial intelligence. Br J Educ Technol. 2023;54(5):1160–73.
2. Bond M, Khosravi H, De Laat M, Bergdahl N, Negrea V, Oxley E, et al. A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. Int J Educ Technol High Educ. 2024;21:4.

3. Stanciu ID, Hernández-García Á, Conde MÁ, Nistor N. Decoding a decade: Trends and evolution in learning analytics: A comprehensive synthesis. Comput Hum Behav. 2025;165:108526.
4. Naeem MR, Khan MA, Khan M, Alruwaili O, Alrashdi I, Alanazi S. Enhancing software quality assurance in ubiquitous learning environments. Comput Hum Behav. 2025;163:108447.
5. Williams J. Stand out of our light: Freedom and resistance in the attention economy. Cambridge University Press; 2018.
6. Hobfoll SE. The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory. Appl Psychol. 2001;50:337–421.
7. Mayer RE. Evidence-based principles for how to design effective instructional videos. J Appl Res Mem Cogn. 2021;10(2):229–40.
8. Pluess M. Individual differences in environmental sensitivity. Child Dev Perspect. 2015;9(3):138–43.
9. Greven CU, Lionetti F, Booth C, Aron EN, Fox E, Schendan HE, et al. Sensory processing sensitivity in the context of environmental sensitivity: A critical review and development of research agenda. Neurosci Biobehav Rev. 2019;98:287–305.
10. Madore KP, Khazenzon AM, Backes CW, Jiang J, Uncapher MR, Norcia AM, et al. Memory failure predicted by attention lapsing and media multitasking. Nature. 2020;587:87–91.
11. Ophir E, Nass C, Wagner AD. Cognitive control in media multitaskers. Proc Natl Acad Sci U S A. 2009;106(37):15583–7.
12. Parry DA, Davidson BI, Sewall CJR, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nat Hum Behav. 2021;5:1535–47.
13. Vallacher RR, Van Geert P, Nowak A. The intrinsic dynamics of psychological process. Curr Dir Psychol Sci. 2015;24:58–64.
14. Diamond A. Executive functions. Annu Rev Psychol. 2013;64:135–68.
15. Schunk DH, DiBenedetto MK. Motivation and social cognitive theory. Contemp Educ Psychol. 2020;60:101832.
16. Theobald M. Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. Contemp Educ Psychol. 2021;66:101976.
17. Elhai JD, Yang H, Montag C. Fear of missing out (FOMO): Overview, theoretical underpinnings, and literature review on relations with severity of negative affectivity and problematic technology use. Braz J Psychiatry. 2021;43(2):203–9.
18. Winne PH, Hadwin AF. Studying as self-regulated engagement in learning. In: Metacognition in educational theory and practice. Hacker DJ, Dunlosky J, Graesser AC, editors. Erlbaum; 1998; pp. 277–304.
19. Pintrich PR. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. J Educ Psychol. 2000;92:544–55.
20. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. 2018;50:195–212.

21. Khosravi H, Shum SB, Chen G, Conati C, Tsai YS, Kay J, et al. Explainable Artificial Intelligence in education. Comput Educ Artif Intell. 2022;3:100074.
22. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: ACM. 2016;pp. 785–794.
23. Ragnedda M. The third digital divide: A Weberian approach to digital inequalities. Routledge; 2017.
24. Childers TL, Houston MJ, Heckler SE. Measurement of individual differences in visual versus verbal information processing. J Consum Res. 1985;12:125–34.
25. Derryberry D, Reed MA. Anxiety-related attentional biases and their regulation by attentional control. J Abnorm Psychol. 2002;111(2):225–36.
26. Awh E, Vogel EK, Oh SH. Interactions between attention and working memory. Neuroscience. 2006;139(1):201–8.
27. Pintrich PR, Smith DAF, Garcia T, McKeachie WJ. A manual for the Motivated Strategies for Learning Questionnaire. National Center for Research; 1991.
28. Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. Self-efficacy beliefs of adolescents. Information Age publishing; 2006. pp. 307–37.
29. Zimmerman BJ, Bandura A, Martinez-Pons M. Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. Am Educ Res J. 1992;29(3):663–76.
30. Panadero E. A review of self-regulated learning: Six models and four directions for research. Front Psychol. 2017;8:422.
31. Leary MR. A brief version of the Fear of Negative Evaluation Scale. Pers Soc Psychol Bull. 1983;9(3):371–5.
32. Fang X, Wo J, Lin X. Development of the Chinese college student adjustment Scale. Psychol Behav Res. 2005;3:95–101.
33. Chemers MM, Hu LT, Garcia BF. Academic self-efficacy and first year college student performance and adjustment. J Educ Psychol. 2001;93:55–64.
34. Hansen BE. Sample splitting and threshold estimation. Econometrica. 2000;68:575–603.
35. Spurk D, Hirschi A, Wang M, Valero D, Kauffeld S. Latent profile analysis: A review and how to guide of its application within vocational behavior research. J Vocat Behav. 2020;120:103445.
36. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw. 2012;48:1–36.
37. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
38. Centola D, Becker J, Brackbill D, Baronchelli A. Experimental evidence for tipping points in social convention. Science. 2018;360(6393):1116–9.
39. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol. 1908;18(5):459–82.
40. Pekrun R. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educ Psychol Rev. 2006;18(4):315–41.

41. Guastello SJ, Peressini AF. The relative influence of drivers and empaths on team synchronization. Nonlinear Dyn Psychol Life Sci. 2021;25(3):357–82.
42. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, Botvinick MM. Toward a rational and mechanistic account of mental effort. Annu Rev Neurosci. 2017;40:99–124.
43. Kruglanski AW, Bélanger JJ, Chen X, Köpetz C, Pierro A, Mannetti L. The energetics of motivated cognition: a force-field analysis. Psychol Rev. 2012;119(1):1–20.
44. Curran PG. Methods for the detection of carelessly invalid responses in survey data. J Exp Soc Psychol. 2016;66:4–19.
45. van Deursen AJAM, van der Zeeuw A, de Boer P, Jansen G, van Rompay T. Digital inequalities in the Internet of Things: Differences in attitudes, material access, skills, and usage. Inf Commun Soc. 2021;24(2):258–76.
46. Fernandes D, Villa S, Nicholls S, Haavisto O, Buschek D, Schmidt A, et al. AI makes you smarter but none the wiser: The disconnect between performance and metacognition. Comput Hum Behav. 2026;175:108779.

Figures

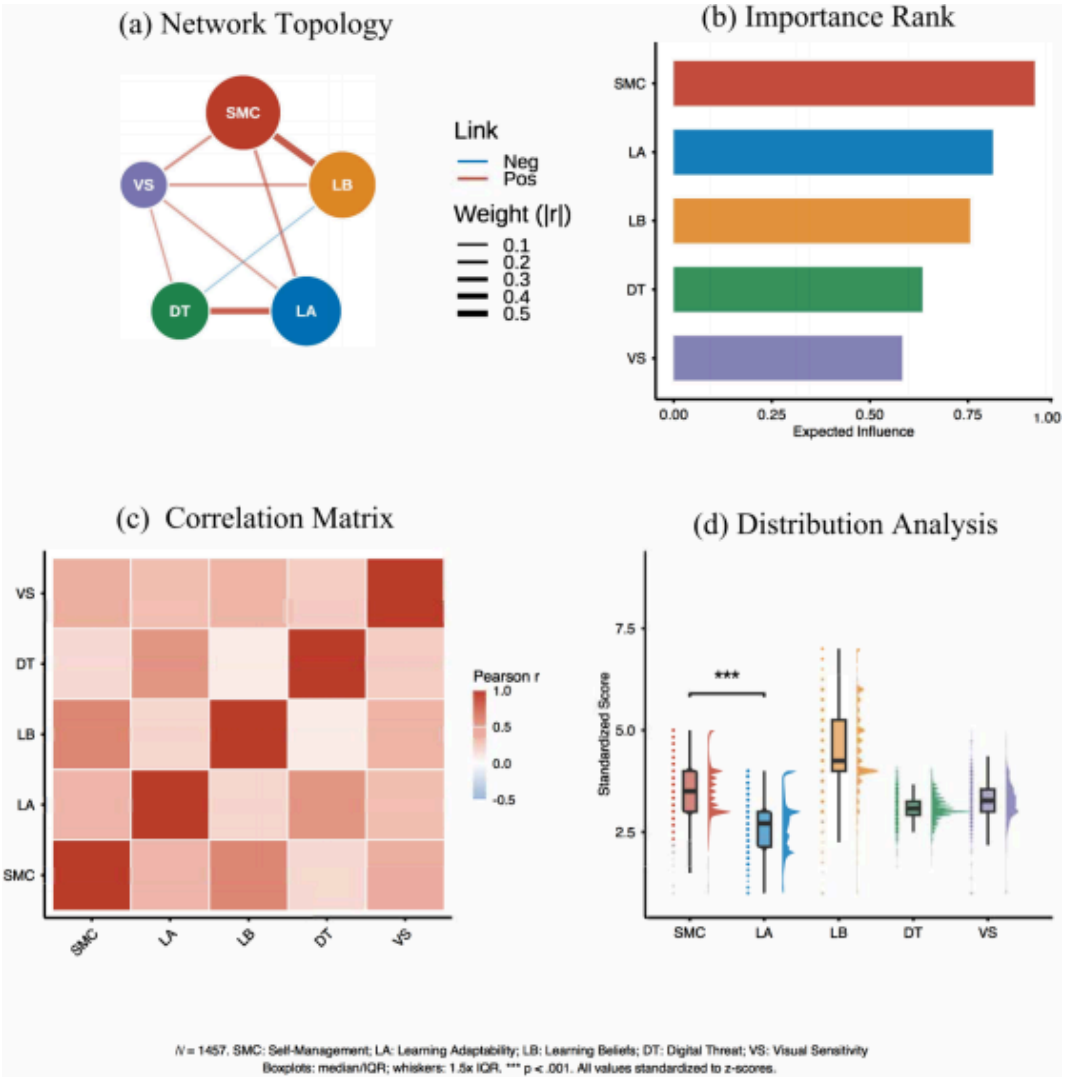

Figure 1

Network psychometric structure of the cognitive system. (a) GGM showing the bifurcated network structure with Threat and Resource clusters. (b) Node centrality indices (expected influence) identifying SMC as the topological hub. (c) Correlation matrix of all study variables. (d) Distribution plots across N = 1,457 participants.

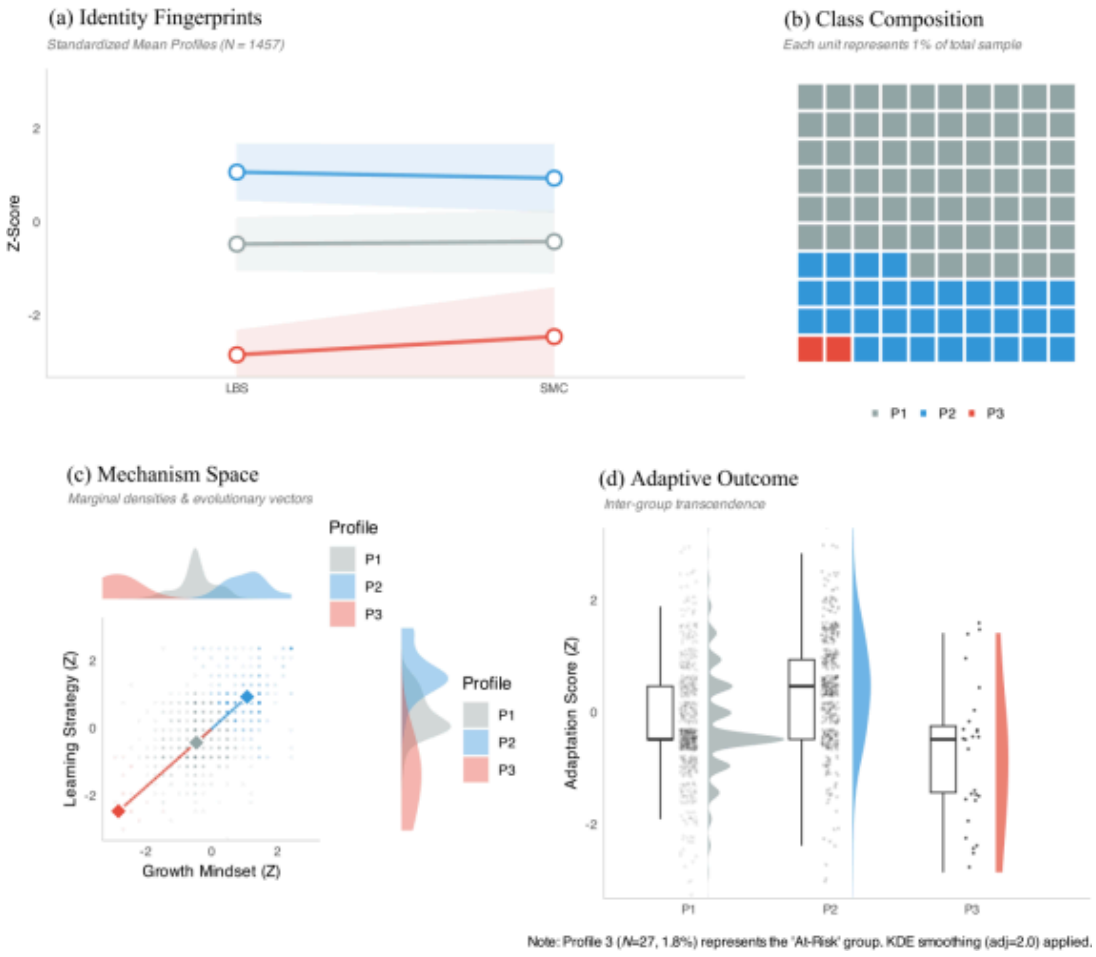

Figure 2

Latent profiles of the Cognitive Firewall. (a) Standardized Z scores of the LBS and SMC across three profiles. (b) Profile composition showing class proportions. (c) Mechanism space analysis illustrating profile vectors. (d) Adaptive outcome distribution confirming separation between profiles.

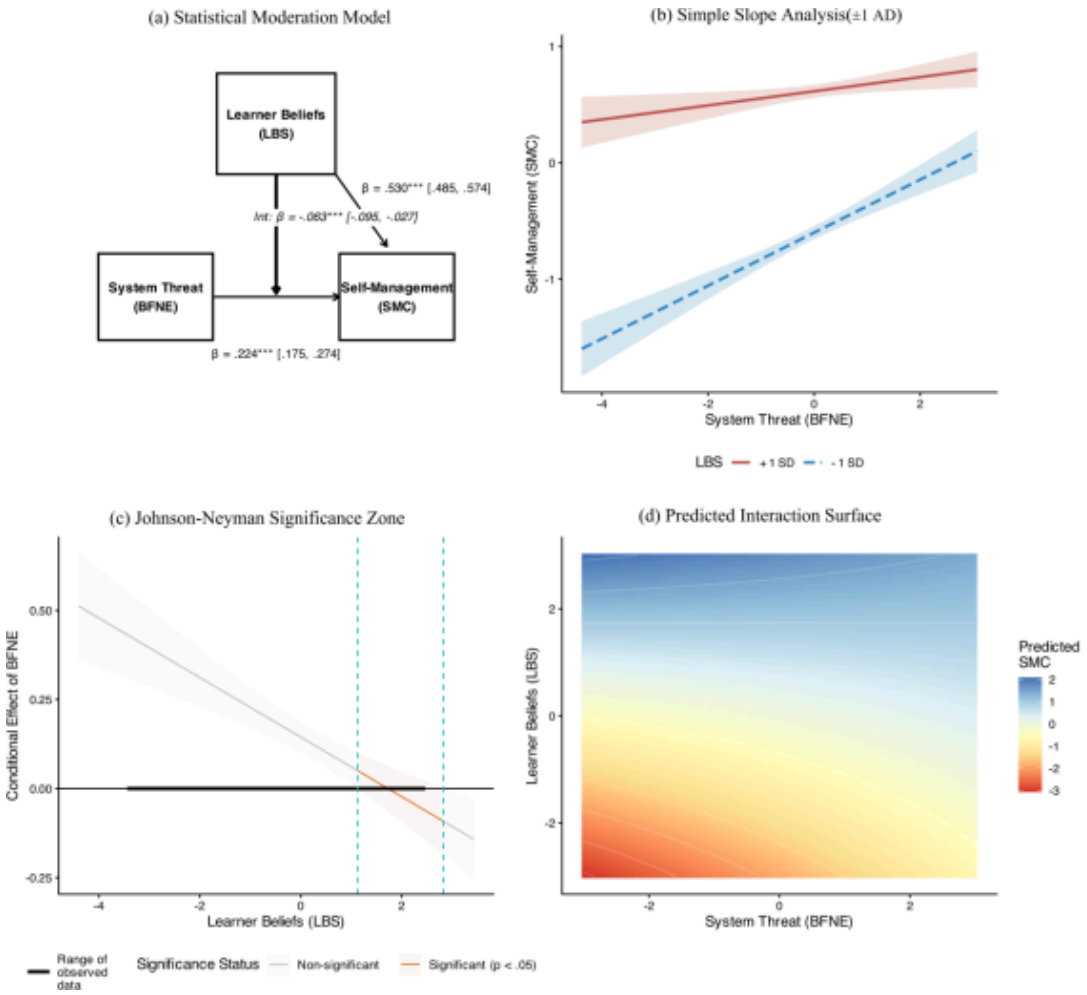

Figure 3

Linear moderated mediation results. (a) Statistical moderation model showing path coefficients. (b) Simple slope analysis showing structural divergence under low vs. high LBS. (c) Johnson–Neyman significance regions. (d) Predicted interaction surface showing stabilization in the high-LBS quadrant.

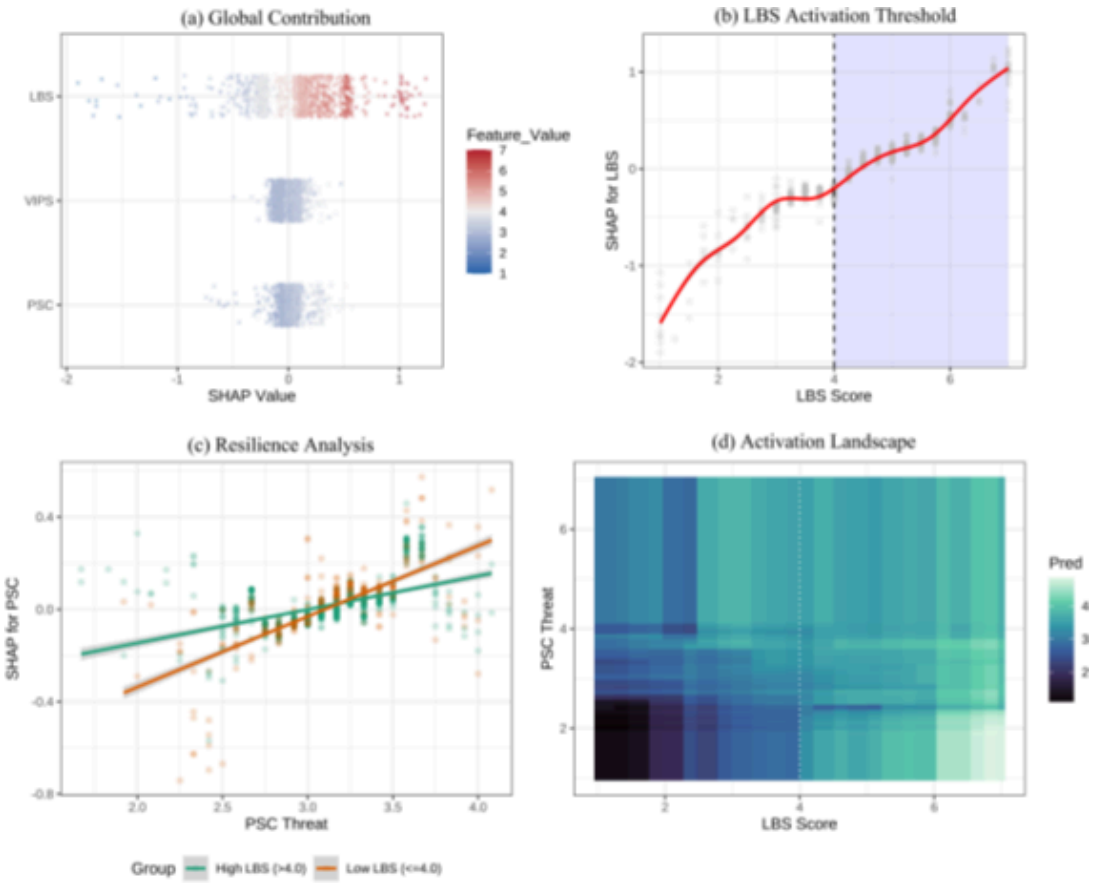

Figure 4

XGBoost and SHAP analysis of nonlinear dynamics. (a) Feature importance hierarchy (Gain). (b) SHAP summary plot. (c) SHAP dependence plot revealing the nonlinear inflection point at LBS ≈ 4.0. (d) Observed vs. predicted values.

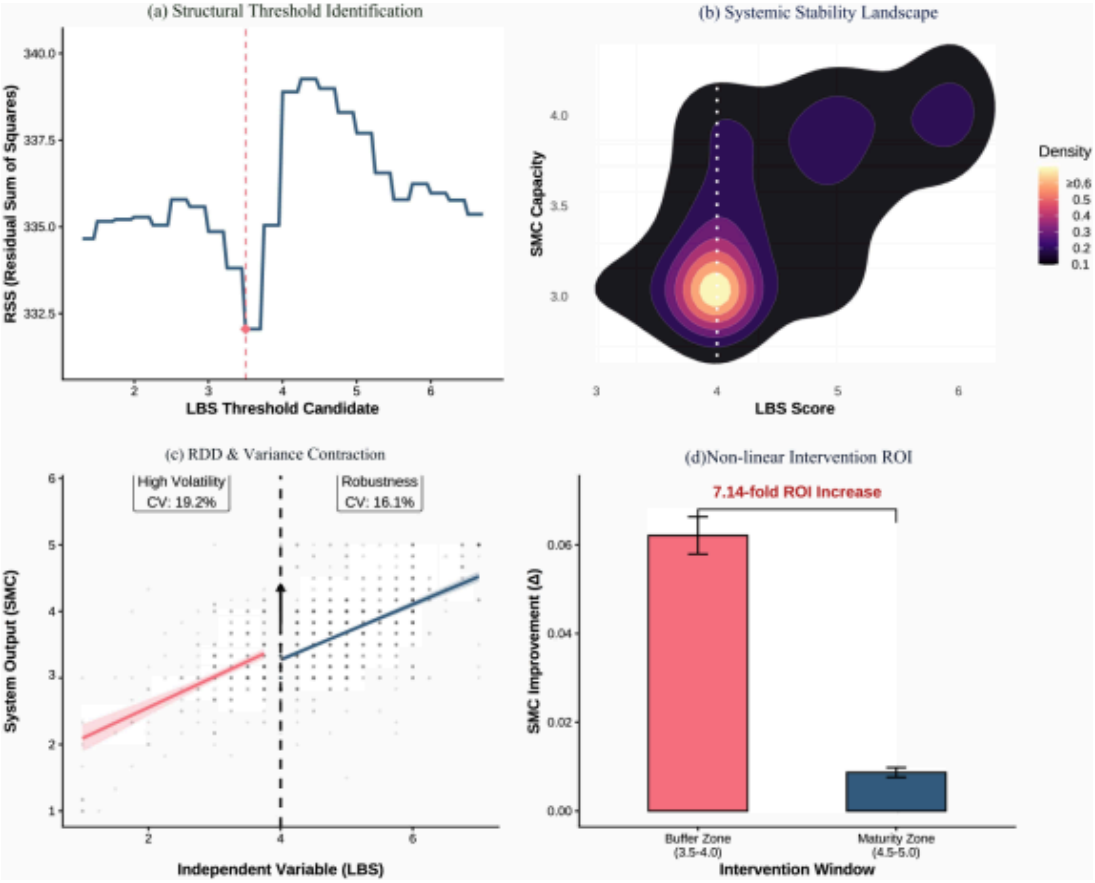

Figure 5

Threshold regression and regime shift validation. (a) LR plot confirming structural breakpoint at LBS = 3.5. (b) Stability landscape showing emergence of the high-density attractor. (c) Regime-specific slopes and buffering effect with variance contraction. (d) Marginal efficiency gain (ROI) across LBS intervals.

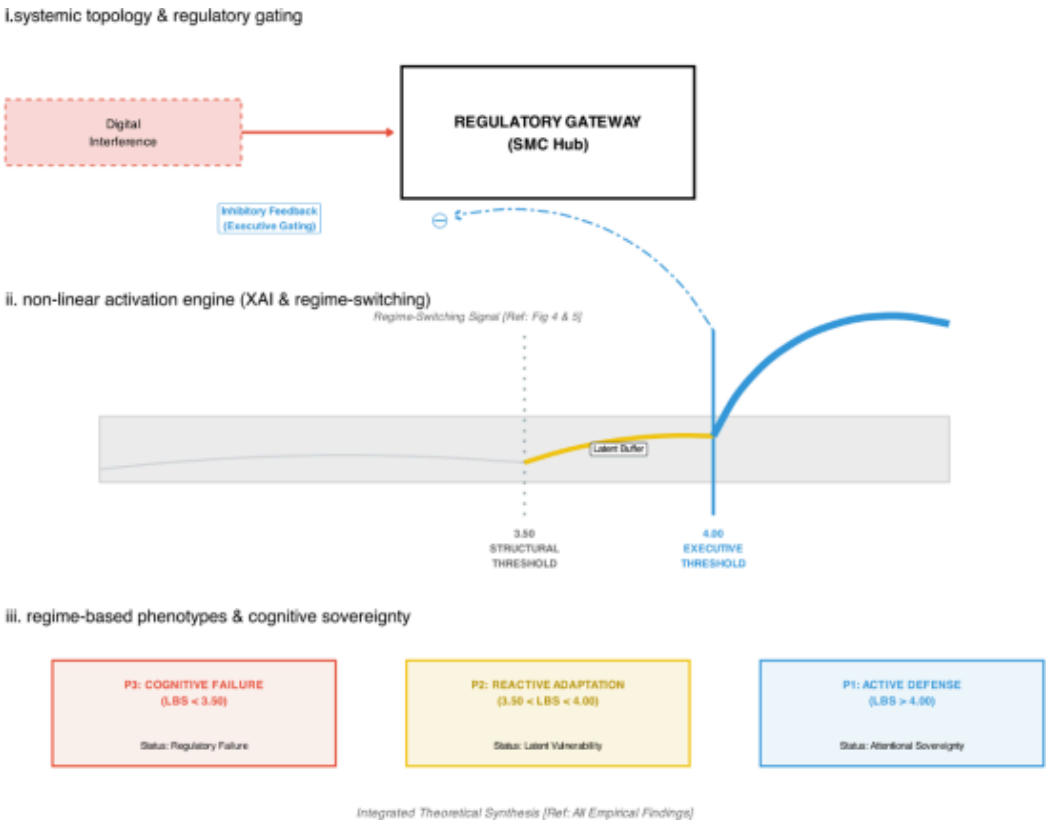

Figure 6

Cognitive Firewall architecture: an integrated synthesis integrating findings from network analysis, XAI, and LPA across three systemic layers.