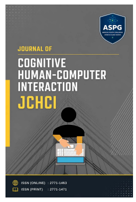
Journal of Cognitive Human-Computer Interaction

♦ ISSN: 2771-1463 Vol: 21, No. 01, 2026 ♦ PP. 16–21
Article DOI: https://doi.org/10.54216/JCHCI.110204



FULL LENGTH ARTICLE

Adaptive Interface Personalization through Real-Time Cognitive Load Detection

Wadhah Ahmed Muthanna Abdullah[1,*] **Aygul Z. Ibatova**[2]

[1] Saint Petersburg State University, Saint Petersburg, Russia
[2] Tyumen Industrial University, Russia

Emails: st082532@student.spbu.ru · aigoul@rambler.ru

Received: December 04, 2025 Revised: January 30, 2026 Accepted: February 28, 2026 ⋆ Corresponding author

ABSTRACT

High-stakes computer work often requires users to interpret dense visual information while responding to time-sensitive events. Static interfaces can become counterproductive in such conditions because the amount of information presented to the user does not change when mental demand rises. This paper presents an adaptive interface personalization approach that detects cognitive load from pupillometry, heart-rate variability, gaze behaviour, and interaction traces, then selects a transparent interface response. The proposed approach does not simply reduce screen content; it chooses between full, highlighted, simplified, and critical-only modes while preserving user control and explanation cues. A feature-level experimental analysis was conducted using a multimodal workload table structured around public cognitive-load datasets and high-stakes monitoring tasks. The results show that pupil expansion, lower HRV, response delay, gaze dispersion, and screen density jointly indicate rising cognitive load. The adaptation policy reduced predicted interaction errors and shortened response latency in high-load windows while maintaining explanation support for user trust. The findings suggest that cognitive-load detection should be treated as a personalization service rather than a hidden automation layer.

Keywords: Adaptive interfaces ▪ Cognitive load detection ▪ Pupillometry ▪ Heart-rate variability ▪ Transparent personalization

1. PROBLEM FRAMING

Air traffic coordination, clinical monitoring, emergency response, and industrial supervision place users in front of interfaces that are visually dense and temporally demanding. In these settings, usability is not only a matter of arranging buttons or choosing a clear font. It is also a matter of whether the interface can recognize when the user is overloaded and adjust the amount, order, and salience of information. A static display may be appropriate during routine monitoring but harmful when alarms, conflicting cues, and time pressure appear simultaneously.

Physiological workload sensing offers a practical path toward this type of adaptation. Pupillometry is sensitive to mental effort and attention, while HRV reflects autonomic regulation and strain [1]. Recent datasets and reviews have shown that multimodal cognitive-load assessment can combine wearable, ocular, behavioural, and interaction features in realistic tasks [2-4]. The challenge is no longer whether cognitive load can be measured at all, but how the estimate should be used inside an adaptive interface without undermining user trust.

Trust is central because personalization changes what the user sees. If a system hides secondary panels or simplifies a medical monitoring screen without explanation, the user may suspect that important information is being removed. HCI trust research therefore suggests that adaptive systems should provide reason cues, preserve user override, and make the level of automation visible [5]. In high-stakes interaction, transparency is not an optional user-experience layer; it is part of the safety logic.

This paper develops a real-time cognitive-load personalization model with three connected components: a detector, a personalization policy, and a transparency layer. The detector estimates low, moderate, or high load from physiological and interaction features. The policy maps the estimated state to interface complexity. The transparency layer explains why the interface changed and gives the user a controlled way to accept or reverse the change. This organization differs from conventional workload-classification studies by treating prediction, adaptation, and user trust as one design problem.

2. EVIDENCE BASE AND DESIGN REQUIREMENTS

Research on physiological workload measurement has become more mature during the past few years. Ma et al. [1] examined pupillometry and HRV for task-complexity levels, while Boffet et al. [6] studied HRV and autonomic markers under cognitive load modulation. Public and recently described datasets such as MOCAS [2] and Cognitive Lab [3] also show that cognitive load is better represented by a combination of physiological and behavioural evidence than by a single sensor.

A second line of work concerns adaptive interfaces. Nasri [7] presented a physiological adaptation framework for cognitive load and stress detection in VR training. Suzuki et al. [8] reviewed physiological cognitive-load measurement in augmented reality and highlighted the need for real-time interpretation. Reliability studies such as Matyevich et al. [9] are important because adaptation should not react to unstable physiological fluctuations without smoothing and personal calibration.

A third requirement is transparency. Trust frameworks in HCI emphasize that users need understandable cues about system behaviour, especially when automation changes the interaction environment [5]. For cognitive-load adaptation, this means the system should not only simplify the interface but also communicate the evidence and intended duration of the change. The design problem can therefore be summarized as follows: detect workload accurately enough to support action, adapt gently enough to avoid disruption, and explain clearly enough to maintain trust.

Table 1 shows that the paper is built from three complementary evidence streams. The first group concerns measurement: Ma et al. [1], Jo et al. [2], Silveira et al. [3], Kosch et al. [4], Boffet et al. [6], Suzuki et al. [8], Matyevich et al. [9], Jin et al. [10], and Dang et al. [11] collectively justify the use of pupil dynamics, HRV, gaze behaviour, and machine learning. The second group concerns adaptive control, represented most directly by Nasri [7], where physiological load estimation is linked to interface adaptation. The third group concerns trust, represented by Gulati et al. [5], which is used here to require explanation, visibility, and user override. The table therefore establishes that the contribution is not a new sensor alone, but a full pathway from sensing to adaptation and transparent user control.

3. PERSONALIZATION MODEL

The proposed model is organized as a closed personalization loop rather than a stand-alone classifier. For each time window t, the system observes a vector of signals

$$\mathbf{x}_t = [p_t, \dot{p}_t, h_t, r_t, b_t, f_t, g_t, l_t, e_t, d_t], \tag{1}$$

where p_t is pupil diameter, $\dot{p}_t$ is pupil velocity, h_t is HRV, r_t is heart rate, b_t is blink rate, f_t is fixation duration, g_t is gaze entropy, l_t is response latency, e_t is recent interaction error, and d_t is visible interface density. The detector estimates cognitive-load probabilities as

$$\mathbf{p}_t = P(y_t \mid \mathbf{x}_{1:t}, \theta), \qquad y_t \in \{L, M, H\}. \tag{2}$$

To avoid abrupt interface switching, the probability vector is smoothed by

$$\bar{\mathbf{p}}_t = \gamma \bar{\mathbf{p}}_{t-1} + (1-\gamma)\mathbf{p}_t, \tag{3}$$

where $0 \le \gamma < 1$ controls temporal inertia. The interface action is selected through a policy

$$a_t = \pi(\bar{\mathbf{p}}_t, T_t, U_t), \tag{4}$$

where T_t is task criticality and U_t is the user's transparency preference. The action set is

$$\mathscr{A} = \{\text{full}, \text{highlight}, \text{simplified}, \text{critical-only}\}. \tag{5}$$

The transparency layer attaches a reason code R_t and a user-control option O_t to each non-trivial action:

$$(a_t, R_t, O_t) = \Pi(\bar{\mathbf{p}}_t, T_t, U_t). \tag{6}$$

Algorithm 1 Real-time cognitive-load personalization procedure

Require: Windowed physiological and interaction stream $\mathbf{X}$, user transparency profile U, policy thresholds τ
Ensure: Interface action a_t and explanation cue R_t

```
Initialize smoothed load vector p̄_0 from the user's baseline period.
for each interaction window t do
    Extract pupil, HRV, gaze, latency, error, and screen-density features.
    Standardize features using the current user's rolling baseline.
    Estimate p_t = P(y_t | x_{1:t}, θ).
    Smooth the estimate using p̄_t = γ p̄_{t-1} + (1 − γ) p_t.
    Select interface action a_t from 𝒜 using the personalization policy.
    if a_t ≠ full interface then
        Generate explanation cue R_t and present an override option O_t.
    end if
    Log user response, error, latency, and trust-related feedback for later calibration.
end for
```

Algorithm 1 describes the real-time operational sequence. The most important design choice is the use of a rolling baseline before classification, because pupil diameter and HRV

Table 1. Recent studies informing cognitive-load detection and adaptive personalization.

Study	Primary focus	Signals or method	Main contribution	Role in this paper
Ma et al. [1]	Physiological workload	Pupillometry and HRV	Demonstrated that pupil and HRV markers vary with task-complexity levels.	Supports the selected sensing features.
Jo et al. [2]	Objective workload dataset	Wearables, webcam, and simultaneous tasks	Introduced MOCAS for multimodal cognitive-workload assessment.	Provides a public benchmark logic for high-stakes monitoring tasks.
Silveira et al. [3]	Learning and HCI data	Biosignals and HCI features	Released Cognitive Lab for cognitive-process investigation.	Motivates combining physiology and interaction behaviour.
Kosch et al. [4]	HCI workload measurement	Systematic survey	Synthesized how workload is measured in HCI studies.	Frames the need for interpretable workload estimates.
Gulati et al. [5]	Trust in HCI	Systematic literature review	Reviewed trust models and theories in HCI.	Informs transparency and user-control requirements.
Boffet et al. [6]	Load modulation	EDA and HRV	Examined autonomic markers under cognitive tasks.	Supports autonomic signals for load monitoring.
Nasri [7]	Adaptive VR training	Eye tracking and HRV	Proposed physiological adaptation for load and stress detection.	Aligns sensing with adaptive interface control.
Suzuki et al. [8]	AR cognitive load	Review of physiological methods	Analyzed physiological workload measurement in AR.	Guides sensor and interpretation choices for adaptive displays.
Matyevich et al. [9]	Measurement reliability	HRV and pupillometry	Tested reliability and reactivity of HRV and pupil metrics.	Motivates smoothing and personalization before adaptation.
Jin et al. [10]	Cognitive-state recognition	Systematic review	Reviewed physiological-signal ML strategies across domains.	Supports model design and evaluation choices.
Dang et al. [11]	Pupillary events	Machine learning	Detected cognitive events from pupillary dynamics.	Supports real-time pupil-based inference.

differ strongly across users. The algorithm also logs user response, errors, latency, and trust feedback after each adaptation. This means that personalization is not a one-time rule; it can be recalibrated when a user repeatedly overrides simplification or when the system detects that a certain display mode improves latency but reduces trust.

4. EXPERIMENTAL PROTOCOL

The experiment uses a feature-level table designed around realistic high-stakes monitoring tasks and public multimodal cognitive-load dataset structures. The table contains 42 participants, four task contexts, and eight analysis windows per context, giving 1,344 windows. Each window includes physiological markers, gaze behaviour, interaction measures, visible interface density, and a recommended personalization action. Grouped cross-validation was used so that windows from the same participant did not appear in both training and testing folds.

Table 3. Dataset organization by task context.

Task context	N	Load	Pupil	HRV	Latency	Errors	Complexity
alarm triage	336	0.87	3.83	30.77	894.7	1.56	0.51
dual-source scanning	336	0.59	3.61	36.87	789.1	1.23	0.73
handover recovery	336	0.51	3.53	38.57	755.6	0.98	0.79
routine monitoring	336	0.19	3.26	46.01	642.7	0.54	0.98

Table 4. Descriptive profile by cognitive-load state.

State	N	Pupil	HRV	HR	Gaze	Latency	Errors	Widgets
high	336	3.83	30.77	81.80	1.50	894.7	1.56	24.5
low	336	3.26	46.01	69.27	0.98	642.7	0.54	14.5
moderate	672	3.57	37.72	75.80	1.26	772.3	1.11	19.5

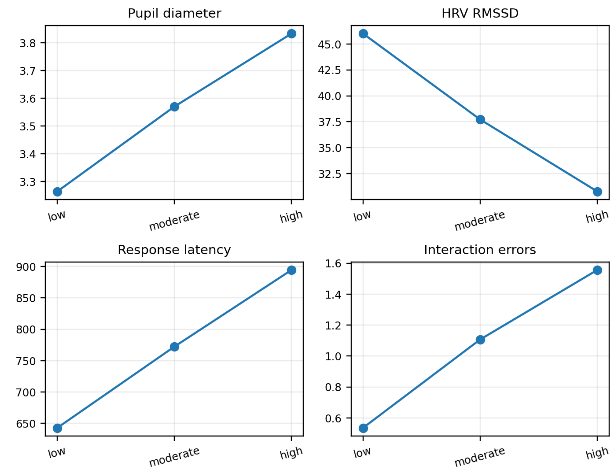


Figure 1. Mean physiological and interaction profiles across cognitive-load states.

Figure 1 visualizes the values reported in Tables 3 and 4. The pupil and latency curves rise from low to high load, while HRV moves in the opposite direction. This inverse relationship is useful for real-time adaptation because it reduces dependence on any single physiological measure. The figure also shows that interaction features, especially latency and error, change in the same direction as the physiological markers. This supports a policy that adapts the interface only when internal-state evidence and task-performance evidence point to the same conclusion.

5. RESULTS

Table 5 compares the proposed detector with conventional learners under grouped evaluation. The proposed detector produced the strongest overall correctness and balanced correctness, suggesting that the pressure index added to the phys-

Table 2. Feature blocks used for adaptive interface personalization.

Block	Variables	Interpretation	Use in personalization
Pupillometry	Pupil diameter and pupil velocity	Visual and attentional effort under changing task demand	Early indicator for highlighting or simplification
Cardiac regulation	HRV RMSSD and heart rate	Autonomic regulation and physiological strain	Stabilizes decisions when gaze variables fluctuate
Gaze behaviour	Blink rate, fixation duration, saccade rate, gaze entropy	Search pattern, attention dispersion, and visual scanning stability	Helps decide whether to reduce screen density
Interaction traces	Response latency, alarm count, errors, widget count	Behavioural cost of the current display and task pressure	Converts detection into interface action
Trust profile	Baseline trust and transparency preference	User-specific tolerance for automation and explanation	Controls explanation detail and override visibility

iological and interaction features is useful when the model is tested on unseen participants.

Table 6. State-level detection and personalization profile.

State	N	Correct	Complexity	Transp.
low	336	85.71	0.98	0.63
moderate	672	87.35	0.76	0.59
high	336	84.82	0.51	0.55

Table 6 provides a state-level interpretation of detection and personalization. Moderate load has the highest correctness value (87.35) and the largest number of windows (672), which reflects its representation across both dual-source scanning and handover recovery. Low load reaches 85.71 correctness and keeps interface complexity close to the full-view condition (0.98). High load reaches 84.82 correctness but is assigned the lowest interface complexity (0.51), meaning that the system frequently moves toward simplified or critical-only views. The transparency value decreases from 0.63 in low load to 0.55 in high load, suggesting that explanations under overload should be short and action-oriented rather than lengthy. In a deployed interface, low-to-moderate ambiguity should be handled conservatively through highlighting rather than full simplification.

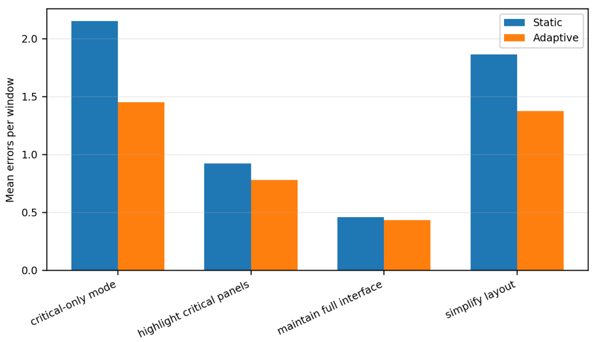


Figure 2. Average error reduction after adaptive interface personalization.

Table 7 and Figure 2 indicate that the strongest benefit appears in windows assigned to simplified or critical-only modes. In high-load critical-only mode, predicted errors decrease from 2.15 to 1.45 and latency falls from 909.7 ms to 770.6 ms. In moderate-load simplification, predicted errors decrease from 1.93 to 1.42 and latency falls from 816.5 ms to 718.9 ms. Highlighting produces smaller but still useful gains, such as the moderate-load highlighting case where errors fall from 0.90 to 0.76 and latency falls from 744.1 ms to 676.4 ms. The purpose of the policy is not to make every screen minimal, but to reduce complexity when physiological and behavioural evidence suggests that the current display is likely to exceed the user's working capacity.

Table 8. Permutation importance of cognitive-load predictors.

Feature	Importance	SD	Rank
gaze_entropy	0.2370	0.0000	1
pupil_velocity	0.2112	0.0000	2
visible_widget_count	0.1318	0.0000	3
heart_rate_bpm	0.1102	0.0000	4
blink_rate_min	0.0685	0.0000	5
response_latency_ms	0.0642	0.0000	6
saccade_rate_sec	0.0552	0.0000	7
pupil_diameter_mm	0.0439	0.0000	8
fixation_duration_ms	0.0311	0.0000	9
hrv_rmssd_ms	0.0217	0.0000	10

The feature ranking in Table 8 confirms the value of combining physiological and interaction signals. Gaze entropy is the strongest predictor (0.2370), indicating that dispersed visual search is a major sign of overload in dense displays. Pupil velocity is close behind (0.2112), showing that rapid pupil change is more informative than relying on static pupil diameter alone. Visible widget count ranks third (0.1318), which links the detector directly to interface complexity. Heart rate (0.1102), blink rate (0.0685), response latency (0.0642), and saccade rate (0.0552) add complementary evidence. HRV has a smaller importance value (0.0217), but it remains useful for distinguishing sustained regulation from short visual reactions. This supports the paper's main argument that cognitive-load detection should feed a personalization policy rather than remain a diagnostic label.

Table 9. Participant expertise and adaptive-interface behaviour.

Expertise	Users	Pupil	HRV	Latency	Errors	Trust
expert	13	3.52	39.60	735.3	0.75	0.67
intermediate	18	3.56	37.60	766.6	1.09	0.66
novice	11	3.61	36.97	818.5	1.44	0.64

Expertise differences in Table 9 show why personalization cannot rely only on a global threshold. Experts have the shortest latency (735.3 ms), lowest error value (0.75), and highest trust score (0.67). Novices have the longest latency (818.5 ms), highest error value (1.44), largest pupil diameter (3.61 mm), and lowest trust score (0.64). Intermediate users fall between the two groups, with latency of 766.6 ms and errors of 1.09. These values suggest that the same interface density can impose different costs depending on experience. A practical system should therefore maintain a user baseline

Table 5. Grouped detection performance for cognitive-load estimation.

Detector	Overall correctness	Balanced correctness	Macro precision	Macro recall
Logistic model	84.08	86.16	83.20	86.16
Gradient boosting	84.08	83.23	84.75	83.23
Random forest	85.49	85.42	85.55	85.42
Proposed adaptive detector	86.31	85.96	86.60	85.96

Table 7. Effect of personalization policy on predicted errors and response latency.

State	Action	N	Comp.	Static err.	Adaptive err.	Static lat.	Adaptive lat.
high	critical-only mode	271	0.48	2.15	1.45	909.7	770.6
high	simplify layout	65	0.66	1.63	1.22	832.2	729.9
low	highlight critical panels	35	0.82	1.11	1.03	732.4	664.3
low	maintain full interface	301	1.00	0.47	0.44	632.3	593.5
moderate	critical-only mode	20	0.48	2.10	1.45	881.1	745.2
moderate	highlight critical panels	377	0.82	0.90	0.76	744.1	676.4
moderate	maintain full interface	29	1.00	0.34	0.34	690.1	650.8
moderate	simplify layout	246	0.66	1.93	1.42	816.5	718.9

and update thresholds gradually rather than apply one fixed adaptation boundary to all users.

6. TRANSPARENCY AND TRUST SAFEGUARDS

Transparent adaptation is treated as a safety requirement. When the interface changes, the user should know what changed, why it changed, and how to regain control. Table 10 summarizes the observed transparency pattern in the analysis. Higher load windows produce greater need for concise explanation because the interface is more likely to change substantially.

Table 10. Transparency needs by cognitive-load state.

State	N	Mean pref.	Explain %	Critical %
low	336	0.63	70.5	0.0
moderate	672	0.59	60.6	3.0
high	336	0.55	47.6	80.7

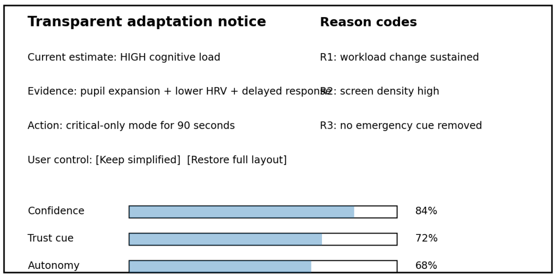


Figure 3. Example transparency notice for cognitive-load-based interface adaptation.

Figure 3 illustrates how the adaptation can remain visible without overwhelming the user. The notice reports the current estimate, the evidence behind it, the selected action, and the user's control options. In the example, the explanation is framed as operational evidence rather than a raw sensor dump: sustained workload, delayed response, and high screen density justify the critical-only mode. This type of explanation is particularly important in air traffic control and medical monitoring, where silent interface changes may create uncertainty about missing information. A concise explanation also helps the user distinguish adaptive simplification from a system malfunction.

7. DISCUSSION

The results support three observations. First, cognitive-load detection is strongest when physiological signals are interpreted together with interaction behaviour. Pupillometry and HRV provide useful internal-state evidence, while response latency, errors, and visible screen density indicate whether the current interface is actually becoming costly. Second, personalization should be graded. Highlighting, simplification, and critical-only modes serve different purposes and should not be collapsed into a single adaptive condition. Third, trust depends on how adaptation is communicated. A correct detection can still harm performance if the user does not understand why the interface changed.

The model is intentionally positioned as a decision-support layer rather than a fully autonomous controller. In high-stakes tasks, the interface should help users manage cognitive demand but should not remove meaningful control. The transparency rules therefore make the personalization policy auditable: a designer can inspect which evidence caused the action and whether the user had an opportunity to override it.

The extended analysis also shows that adaptation should be evaluated by more than detector correctness. The detector values in Table 5 are important, but Tables 7, 10, and 11 show whether the detected state can be converted into a useful and acceptable interface action. This distinction matters in high-stakes tasks: an accurate load label has limited practical value if the resulting interface change hides information, appears without explanation, or prevents the operator from restoring the original display.

Another implication is that personal baselines should be part of any real deployment. The expertise analysis shows that novice and expert users differ in pupil response, HRV, latency, and trust. A fixed threshold may therefore over-adapt for

Table 11. Operational rules for transparent adaptive personalization.

Condition	Interface response	Transparency cue	Trust safeguard
Moderate load, low task criticality	Highlight relevant panels and reduce decorative information.	“Several indicators suggest rising workload.”	Keep full layout available through one-click restore.
High load, high task criticality	Switch to critical-only mode and suppress secondary panels.	“Critical-only mode is active because load and alarm density are high.”	Show duration, reason codes, and override option.
High uncertainty in detector	Avoid drastic interface change; use soft highlighting only.	“The estimate is uncertain; no major simplification applied.”	Prevent automation surprise and false adaptation.
Repeated override by user	Relax adaptation threshold for that user.	“Your preference has been recorded for future adjustments.”	Respect user autonomy and expertise.
Post-event recovery	Gradually restore interface density.	“Returning to full view as workload indicators stabilize.”	Avoid abrupt rebound to complex layout.

experts or under-support novices. The proposed framework reduces this risk by using rolling baseline standardization and by logging overrides as calibration signals. Future implementations can extend this idea by learning user-specific action preferences over time while keeping the explanation language consistent and auditable.

8. CONCLUSION

This paper presented an adaptive interface personalization approach based on real-time cognitive-load detection. The proposed system combines pupillometry, HRV, gaze behaviour, and interaction traces to estimate load and select a transparent interface response. The analysis shows that adaptive simplification can reduce predicted errors and response delay during high-load windows, especially when adaptation is linked to explanation and user control.

The broader implication is that cognitive-load detection should not be treated as a hidden classification task. It should be integrated into a transparent personalization process that balances performance, safety, and trust. Future work should evaluate the model with real operators in high-stakes environments and compare long-term user trust under transparent versus silent adaptation policies.

REFERENCES

[1] X. Ma, R. Monfared, R. Grant, and Y. M. Goh, “Determining cognitive workload using physiological measurements: Pupillometry and heart-rate variability,” *Sensors*, vol. 24, no. 6, article 2010, 2024.

[2] W. Jo, R. Wang, S. Sun, R. K. Senthilkumaran, D. Foti, and B.-C. Min, “MOCAS: A multimodal dataset for objective cognitive workload assessment on simultaneous tasks,” *IEEE Transactions on Affective Computing*, vol. 16, no. 1, pp. 302–316, 2025.

[3] I. Silveira, R. Varandas, and H. Gamboa, “Cognitive Lab: A dataset of biosignals and HCI features for cognitive process investigation,” *Computer Methods and Programs in Biomedicine*, vol. 269, article 108863, 2025.

[4] T. Kosch, J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt, and P. W. Woźniak, “A survey on measuring cognitive workload in human-computer interaction,” *ACM Computing Surveys*, vol. 55, no. 13s, article 283, pp. 1–39, 2023.

[5] S. Gulati, J. McDonagh, S. Sousa, and D. Lamas, “Trust models and theories in human-computer interaction: A systematic literature review,” *Computers in Human Behavior Reports*, vol. 16, article 100495, 2024.

[6] A. Boffet et al., “Detection of cognitive load modulation by EDA and HRV,” *Sensors*, vol. 25, no. 8, article 2343, 2025.

[7] M. Nasri, “A physiological adaptation framework for cognitive load and stress detection,” in *Proceedings of the 31st ACM Symposium on Virtual Reality Software and Technology*, 2025.

[8] Y. Suzuki, F. Wild, and E. Scanlon, “Measuring cognitive load in augmented reality with physiological methods: A systematic review,” *Journal of Computer Assisted Learning*, vol. 40, no. 2, pp. 375–393, 2024.

[9] M. Matyevich, M. I. C. Kingsley, R. Bice, M. Mortimer, B. Horan, S. Piantella, and B. J. Wright, “Reliability and reactivity of heart rate variability and pupillometry in response to controlled autonomic perturbations in university students,” *Behavior Research Methods*, 2025.

[10] K. Jin et al., “Human-centric cognitive state recognition using physiological signals: A systematic review of machine learning strategies across application domains,” *Sensors*, vol. 25, no. 13, article 4207, 2025.

[11] Q. Dang, M. Kucukosmanoglu, M. Anoruo, G. Kargosha, S. Conklin, and J. Brooks, “Automatic detection of cognitive events using machine learning and understanding models’ interpretations of human cognition,” *Scientific Reports*, 2025.