Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 51–62
https://doi.org/10.54941/ahfe1007064

Seeing the Invisible Load: XR + Multimodal Sensing for Cognitive Ergonomics in Industrial Training

Jessica M. Johnson[1] and Andwele Grant[2]

[1]Old Dominion University, Norfolk, VA 23435, USA
[2]Commonwealth Center for Advanced Manufacturing, Disputanta, VA 23842, USA

ABSTRACT

Extended reality (XR) technologies are increasingly positioned as disruptive Industry 5.0 tools for human-centric industrial training and intelligent human–system integration. Coupled with multimodal sensing (eye tracking, EEG, HRV, GSR, and other physiological signals), XR environments promise to make otherwise invisible cognitive demands observable, especially for novice trainees entering complex industrial settings. Yet the evidence base is fragmented: (1) there is no quantitative synthesis of the cognitive ergonomics benefits of XR plus sensing; (2) little is known about which XR–sensor configurations yield the strongest effects; (3) prior reviews rarely focus on industrial and manufacturing tasks; (4) multimodal signals are used predominantly for post-hoc diagnosis rather than real-time adaptation; and (5) trade-offs between egocentric, in-situ capture and controlled laboratory configurations are poorly characterized. This paper presents a meta-analysis of empirical studies that (a) used XR for training or performance support, (b) integrated at least one multimodal sensing channel, and (c) reported training- or work-relevant outcomes such as workload, situation awareness, task performance, or transfer. For this paper, we focused on manufacturing and industrial tasks (e.g., assembly, inspection, maintenance) and on novice or early-career operators. The synthesis yields evidence indicating where XR plus multimodal sensing robustly improves cognitive ergonomics for industrial novices, where effects are weak or inconsistent, and which broad modality–sensor pairings are most often associated with reduced workload, enhanced situation awareness, and lower error rates. Results indicated that egocentric, in-situ capture increases ecological validity without systematically degrading training and performance benefits but also reveal a major gap: most systems treat multimodal data as diagnostic rather than as inputs to intelligent, closed-loop adaptation. Building upon these findings, we design guidance for intelligent interfaces and human-machine teaming in industrial systems emphasizing XR modalities and adaptive policies for future cognitive ergonomics-Industry 5.0 embedded systems.

Keywords: Hybrid cognitive-centered systems, XR training, Cognitive ergonomics, Human-machine teaming, Adaptive cognitive systems

INTRODUCTION

Industrial training increasingly resembles a hybrid cognitive-centered system: a cognitive ecosystem where people coordinate with representations,

Received November 25, 2025; Revised December 23, 2025; Accepted January 7, 2026; Available online February 1, 2026

interactive tools, and decision-support features that function as artificial cognitive components. XR is a natural substrate for this shift because it can externalize task structure, guide attention through spatial and interface constraints, and enable safe rehearsal of complex work, aligning with Industry 5.0's emphasis on worker well-being, adaptability, and human–technology collaboration (Nahavandi, 2019; Gladysz et al., 2023). Yet XR alone often misses the "invisible layer" that determines whether training accelerates expertise or quietly engrains brittle habits: cognitive load, attentional stability, and physiological stress responses under real industrial constraints. Multimodal sensing addresses this gap by translating latent cognitive states into measurable indicators of workload, attentional control, and performance-relevant strain (Khan et al., 2025; Zakeri et al., 2022).

Wearable sensing channels provide complementary windows into these dynamics and can function as cognitive state estimators when interpreted appropriately. Eye tracking captures attentional allocation and oculomotor proxies of load, EEG supports workload classification and cognitive control tracking, and HRV/EDA index autonomic activation linked to effort and stress. However, synthesis work emphasizes that multimodal validity hinges on task design, preprocessing, and interpretation models, even when combined sensors increase sensitivity to load differences (Khan et al., 2025). In industrial human–robot interaction, EEG-based workload assessment has demonstrated feasibility for distinguishing workload conditions in simulated tasks, supporting the claim that physiological sensing can index cognitive ergonomics risks in complex collaboration settings (Zakeri et al., 2022).

The core question, then, is not whether sensors can be added to XR, but whether XR + sensing is integrated as a genuine human–artificial cognitive system. In such systems, the interface and its logic operate as a teammate by shaping information flow, supporting decision-making, and stabilizing coordination through guidance overlays, sequencing, attention cues, error detection, and pacing. Yet a persistent gap is that sensing is used mainly for retrospective diagnosis rather than real-time adaptation, limiting its capacity to prevent overload and attentional collapse during learning. Evidence syntheses further indicate that outcomes depend not only on XR use, but on how interaction design and feedback timing structure cognition in the moment, underscoring that the interface is part of the cognitive machinery of training (Bödding et al., 2025).

METHOD

Protocol, Search Strategy, and Eligibility Criteria

A written protocol specified the review's aims, eligibility criteria, outcomes, moderators, coding instructions, and the analytic strategy, including effect-size computation, random-effects pooling, heterogeneity assessment, and bias/sensitivity checks (Page et al., 2021). The scope targeted industrial and manufacturing training or performance-support use cases in which XR functioned as an intelligent interface within a broader human-technology system, and was paired with at least one wearable or physiological sensing modality capable of indexing cognitive state (e.g., eye-tracking, EEG, HR/HRV, EDA/GSR), consistent with the multimodal workload literature (Khan et al., 2025; Zakeri et al., 2022). Searches were conducted across scholarly databases commonly used for XR and human-systems research (e.g., Scopus/Web of Science, IEEE Xplore, ACM Digital Library), supplemented by backward and forward citation chaining of eligible articles and relevant synthesis papers from January 2015 through August 2025. Boolean operators, truncation, and terminology variants were used to maximize coverage across venues and disciplines.

Eligibility criteria required peer-reviewed empirical studies (journal articles or peer-reviewed conference proceedings) using human-subjects experimental or experimental designs that evaluated an XR condition in an industrially relevant task. Studies had to integrate at least one multimodal sensing channel during XR use and report inferential statistics and/or sufficient quantitative information to compute or derive standardized effects (e.g., means/SD and sample sizes, or convertible t, F, p values). Included studies reported at least one of four target outcome families: 1) workload, 2) situation awareness, 3) task performance, and/or 4) transfer (near/far transfer; retention or post-training performance under changed conditions).

Study Selection

The database search returned 1,792 records. After deduplication and title/abstract screening for relevance to adaptive or adjustable function allocation, approximately 82 full-text articles were assessed in detail. About 40 articles were excluded at the full-text stage because they lacked human data or did not provide sufficient statistical information for effect-size estimation. Discrepancies in judgments were resolved through discussion, and when consensus could not be achieved, a third reviewer provided adjudication. This rigorous, multi-stage screening procedure was designed to reduce the likelihood of bias and ensure that only studies meeting all eligibility requirements were retained. This procedure minimized selection bias and ensured that only studies meeting all eligibility requirements were retained for synthesis. Ultimately 42 studies (total N ≈ 1,200 participants) were deemed suitable for meta-syntheses.

Data Extraction and Coding

For each included study, we extracted bibliographic information, sample characteristics (including trainee expertise when reported), task domain (e.g., assembly, inspection, maintenance, safety/monitoring, human–robot collaboration), XR modality (AR/VR/MR), and sensing configuration (eye tracking, EEG, HR/HRV, EDA/GSR, and multi-sensor combinations). Given known sensitivity and validity constraints in multimodal workload measurement, we also coded implementation details that could affect interpretability, including sensor placement/form factor (e.g., integrated HMD eye tracking vs. external eye tracking; wearable EEG vs. laboratory

EEG), measurement timing (continuous vs. segmented), and whether sensing was used diagnostically (post-hoc) or as part of a closed-loop adaptive policy that altered guidance, pacing, or interface state during training.

Outcomes were organized into four pre-specified families to reduce construct proliferation. Workload included validated multi-item instruments (e.g., NASA-TLX or comparable scales) and explicitly labelled mental workload measures. Situation awareness included validated SA scales, SA probe methods, or task-embedded SA proxies aligned with authors' operational definitions. Task performance included accuracy, time, error rate, throughput, completion quality, and comparable objective metrics. Transfer included delayed tests, retention assessments, and near/far transfer tasks (including performance without scaffolds, changes in context/conditions, or novel variants of the trained task). When multiple indicators were reported within the same outcome family at the same timepoint, we prioritized validated primary endpoints if authors specified them; otherwise, we computed within-study composite (e.g., averaging standardized indicators) to avoid overweighting single studies. Repeated-measures designs were coded at the authors' primary post-test/end-of-training timepoint; alternative timepoints (e.g., mid-training, retention) were retained for sensitivity analyses. Effect sizes were extracted or computed for each eligible comparison. The primary standardized metric was Hedges' *g* (small-sample corrected standardized mean difference), directionally aligned so that positive values reflected beneficial effects of XR + sensing (e.g., lower workload, higher SA, better performance, better transfer). For outcomes where lower scores indicate improvement (e.g., workload, errors, time), signs were reversed accordingly to maintain consistent interpretation.

Evidence Synthesis

Effect sizes were synthesized using random-effects models to account for true effect variability across tasks, XR modalities, sensing suites, and training designs. Pooled estimates are reported with 95% confidence intervals, alongside standard heterogeneity statistics (Cochran's Q and I^2) to quantify between-study variability. Moderator patterns were examined to characterize when XR + sensing operates most effectively as a hybrid cognitive-centered system. Planned moderators included XR modality (AR vs. VR vs. MR), sensing suite (single modality vs. multimodal), adaptive policy type (diagnostic-only vs. closed-loop adaptation), task class (assembly/maintenance/inspection/safety-monitoring/human–robot collaboration), training setting (laboratory vs. in-situ/field when reported), and trainee expertise (novice/early-career vs. mixed samples). Where cell sizes supported inference, subgroup analyses and meta-regression were used to explore moderation; otherwise, patterns were summarized descriptively with attention to confidence intervals and heterogeneity shifts.

To address dependence from multiple effects contributed by the same study (e.g., multiple outcomes or multiple comparisons), primary models were complemented with robustness checks using (a) within-study averaging to yield one independent effect per study per outcome family and

(b) cluster-robust variance estimation to adjust standard errors for study-level clustering. Influence was examined via leave-one-out diagnostics and outlier inspection; sensitivity analyses evaluated whether conclusions changed when conspicuous outliers or high-risk-of-bias designs were excluded. Potential small-study effects and publication bias were assessed visually using funnel plots and analytically using Egger-type regression tests when feasible; trim-and-fill was used, if at all, only as a sensitivity illustration given the likelihood of genuine heterogeneity in XR training research. All analyses were conducted in R using established meta-analytic workflows

RESULTS

Workload Effects in Hybrid Cognitive-Centered Systems

Most studies were conducted as controlled experiments in laboratory or training environment settings, with a smaller subset evaluating XR in real operational environments. Across the evidence base, VR ($k = 20$), AR ($k = 15$), and MR/hybrid systems ($k = 7$) were represented, spanning industrial domains such as manufacturing assembly, maintenance and troubleshooting, construction and safety, and operational readiness contexts. Sensor integration was common but uneven: 62% of studies (26/42) incorporated at least one sensing modality, most frequently eye tracking ($k = 18$) and EEG ($k = 10$), followed by HR/HRV ($k = 7$) and EDA/GSR ($k = 5$), with NASA-TLX and performance measures still prevalent in many designs. Critically, only three studies implemented closed-loop adaptation where sensor inputs influenced the training experience in real time, and all of these were laboratory prototypes, underscoring a major translational gap between multimodal measurement and adaptive system control. Qualitatively, trainees generally reported high engagement and preference for XR after initial acclimation, but several studies also documented early-session overload when interfaces were dense or interaction mechanics were unfamiliar, reinforcing XR's dual potential to either reduce workload through well-designed guidance or increase it through added complexity. XR-based training demonstrated a significant overall benefit for task performance across studies, with a pooled random-effects estimate of Hedges' $g = 0.67$ (95% CI [0.50, 0.85]) and moderate heterogeneity ($I^2 = 55\%$). Performance gains were broadly comparable across AR and VR, while task type moderated effects, with stronger benefits for procedural and motor-intensive tasks than for primarily cognitive or decision-focused scenarios. Importantly, studies including sensors did not, on average, show larger performance gains than those without sensors, consistent with the observation that sensing was predominantly used diagnostically rather than as an adaptive mechanism during training.

Situation Awareness Under Complex Human-Systems Integration

Situation awareness (SA) was measured less often and less consistently than workload or performance, limiting pooled estimation. SA improvements were most likely when XR operates as an information-shaping interface rather than a realism-maximizing display. Systems that manage information density, reduce clutter, and guide attention tend to stabilize novice perception–interpretation–projection loops, with eye tracking and pupillometry frequently revealing whether trainees maintained broad scan coverage versus collapsing into attentional tunneling (Souchet et al., 2022; Gualtieri et al., 2022). Workload results contextualize this: across 22 studies using NASA-TLX (or similar), subjective workload did not differ overall from control (pooled $g = -0.05$, 95% CI [−0.20, 0.10]), masking a split where AR tended to lower workload ($g = -0.25$) and VR trended slightly higher ($g = 0.15$) without a reliable subgroup difference. Physiological signals often exposed "hidden" strain during SA-relevant moments, with lower HRV in XR vs. control in five studies and EDA spikes aligning with confusion points, suggesting that SA vulnerability in XR is driven more by interface timing and attentional control than by immersion per se.

Task Performance Effects and Error Reduction

Task performance indicated the strongest meta-analytic support. Across 30 studies, XR produced a medium-to-large, pooled advantage over non-XR controls (Hedges' $g = 0.67$, 95% CI [0.50, 0.85]) with moderate heterogeneity ($I^2 = 55\%$), reflecting faster completion, higher accuracy, and fewer errors/assistance needs in common industrial tasks. Effects were comparable for AR and VR (both roughly $g \sim 0.6$–0.7), but task type moderated outcomes: procedural motor tasks yielded larger gains ($g \approx 0.80$) than primarily cognitive/decision tasks ($g \approx 0.40$). Studies adding sensors did not show larger performance gains on average, consistent with sensing being mostly diagnostic rather than adaptive; the performance mechanism appears to be XR's ability to structure cognition through well-timed cues and error-preventive scaffolds. Illustratively, AR step overlays reduced search and memory demands and improved efficiency while shifting gaze toward the workpiece (Limbu et al., 2018).

Transfer as a Test of Adaptive Cognitive Systems

Transfer was the sparsest outcome category, constraining quantitative synthesis, but available evidence suggests transfer is most likely when XR + sensing supports strategy learning (attention allocation, verification routines, decision cueing) rather than step-following alone. This gap is amplified by the rarity of closed-loop designs: only three studies implemented real-time sensor-driven adaptation, all in laboratories (e.g., EEG-informed VR difficulty adjustment; Dey et al., 2019). Where transfer was measured, multimodal patterns suggested that higher cognitive demand during XR learning can translate into more efficient performance later: Mondellini et al. (2024) reported higher neural workload during VR training but more efficient neural effort during subsequent testing. In safety training, combining EEG and GSR detected overwhelmed states at ~85% accuracy and coincided with higher delayed knowledge outcomes (~15% retention gain) despite slightly higher perceived demand (Zakeri et al., 2023). Overall, transfer functions as the clearest litmus test for whether XR + sensing has matured into an adaptive cognitive system; the current literature is promising but underpowered and strongly points to the need for field-valid, closed-loop trials.

DISCUSSION

Human-Machine Teaming and Collaboration: Why "Diagnostic-Only" Is Not Enough

A consistent gap across the included studies is that multimodal sensing is used primarily to *observe* cognition rather than to *shape* it. Many XR systems utilized in the current research synthesis reached instrumentation but stopped short of integration: they can *measure* attention, arousal, or workload, yet they do not convert those signals into teammate-like behaviors in the interface or training logic. This distinction matters because the central promise of XR + sensing is to reveal and manage invisible cognitive load: the latent strain, attentional drift, and physiological activation that can accumulate even when outward performance looks acceptable (Khan et al., 2025; Zakeri et al., 2022). When sensing remains diagnostic-only, invisible load becomes merely documented, not engineered, leaving the training system unable to intervene when cognitive conditions are most fragile. This is particularly consequential for human-machine teaming in industrial training because some of the most damaging training outcomes are "quiet failures" that masquerade as success. Novices can complete a procedure while still learning the wrong thing: they may tunnel attention onto salient but non-diagnostic cues, build shallow situation models, or become dependent on prompts rather than learning verification routines and decision cues. These failures often do not appear as immediate errors, but they show up as brittle performance under variation, poor transfer, and mis calibrated trust in automation. Multimodal sensing is uniquely positioned to expose these hidden dynamics. Eye tracking can reveal whether the trainee is scanning the environment in a way that supports situation awareness or whether gaze collapses into instruction-following loops; pupillometry can flag rising cognitive effort as task complexity increases; EEG and autonomic measures can detect sustained or spiking load that is not verbally reported (Khan et al., 2025; Souchet et al., 2022). In this sense, "invisible cognitive load" is not a metaphor, but a measurable phenomenon: physiological and behavioral signatures can indicate when a trainee is coping through compensatory effort rather than developing stable, transferable skill.

However, measurement alone does not change outcomes. The literature's limited number of closed-loop systems illustrates the core issue: without adaptation, the sensing layer functions like a post-game sports replay, not like a teammate making real-time calls. The evidence base shows that XR outcomes are sensitive to interaction design and feedback timing, implying that the interface itself is part of the cognitive machinery of training

(Bödding et al., 2025). If invisible load spikes at moments of confusion or overload, a diagnostic-only system merely records it; an integrated human–artificial cognitive system would respond by adjusting information density, sequencing guidance, or triggering micro-coaching at the moment the trainee's cognitive control is most taxed. That requires explicit adaptive policies: rules or learned mappings from sensed state (e.g., overload probability, attentional instability, fatigue indicators) to interface actions (e.g., reduce overlay clutter, highlight the next discriminative cue, slow pacing, introduce a confirmatory checklist, or temporarily shift from hints to reflective prompts). Even relatively simple policies may be meaningful if they prevent trainees from practicing under persistent invisible load that encourages shortcuts and dependency.

The scarcity of real-time adaptation in the included studies suggests that the field has not yet fully embraced the human-machine teaming implications of XR + sensing. Industry 5.0 framing emphasizes human-centric integration and worker well-being, but human-centricity in training must include the cognitive layer: systems should protect attentional stability and regulate workload while still supporting learning and autonomy (Nahavandi, 2019; Gladysz et al., 2023). In practice, this implies a shift in evaluation criteria. Instead of asking only whether XR improves immediate task performance, an IHSI-aligned program should ask whether the system detects and mitigates invisible cognitive load in ways that preserve situation awareness, prevent attentional tunnelling, and cultivate transfer-ready strategies. Without that shift, XR risks producing trainees who look competent inside the simulator but are fragile in real work, precisely because the most important cognitive dynamics remained invisible to the training logic even if they were visible in the data logs.

Design Guidance for Complex Evolutionary and Adaptive Cognitive Systems in Industry 5.0 Systems

The value of multimodal sensing is realized only when it is treated as a control input to an intelligent teammate-like interface rather than as an after-action metric. The next generation of XR training systems should therefore operationalize invisible cognitive load as an actionable state: sensed, interpreted, and used to trigger adaptive support that is timely, minimal, and transparent. That is the difference between a system that merely measures cognition and one that engineers cognitive readiness through human-machine teaming (Antonaci et al., 2024; Ayres et al., 2021; Gualtieri et al., 2022; Souchet et al., 2022; Zakeri et al., 2023). To translate these findings into implementable design choices, Table 1, below, summarizes five evidence-informed principles, each framed around the type of invisible cognitive load it targets, the minimal sensing needed to detect it, and the corresponding adaptive policy patterns that enable intelligent, human-centered control in Industry 5.0 training systems.

Table 1: Design guidance for adaptive XR + sensing systems to detect and regulate invisible cognitive load for Industry 5.0.

Design Guidance	Invisible Cognitive Load Target	Minimal Sensing Needed (Examples)	Adaptive Policy Examples (Teammate-Like Interface)
Design XR + sensing as a hybrid cognitive-centered system: specify the cognitive functions to stabilize first, then pick XR representations and sensing channels (Antonaci et al., 2024; Gualtieri et al., 2022).	Invisible overload from novelty, multitasking, or poor cueing; attentional instability that still permits task completion.	Eye tracking + pupillometry for attention/effort; optional EEG/EDA when signal quality is feasible (Souchet et al., 2022; Zakeri et al., 2023).	Define control objectives (attention control, workload regulation, SA stability) and align overlays, prompts, and task structure to those objectives. *Primary outcomes to evaluate:* Workload (subjective + physiology), SA, task performance, transfer.
Prioritize adaptive policies over bigger sensor stacks: fewer signals + a clear action rule beats many signals used only for dashboards (Ayres et al., 2021; Dey et al., 2019).	Effort spikes and fatigue that do not immediately appear as errors ("hidden" strain).	One robust channel can suffice (e.g., gaze/pupil), with optional HRV/EDA for arousal context (Souchet et al., 2022; Zakeri et al., 2023).	Trigger rules: reduce information density when overload risk rises; adjust pacing/difficulty; switch feedback modes during fatigue. *Primary outcomes to evaluate:* Reduced invisible load without performance loss; fewer errors; stable attention patterns.
Use sensing to protect SA, not to increase alerts: adapt by subtracting clutter and sequencing information (Gualtieri et al., 2022; Souchet et al., 2022).	Attentional tunnelling and clutter-driven SA erosion (augmentation paradox).	Eye tracking for scan coverage/tunnelling; pupil dilation for effort spikes (Souchet et al., 2022).	Clutter suppression; cue prioritization at decision points; adaptive highlighting of discriminative cues; delay nonessential prompts under overload. *Primary outcomes to evaluate:* SA stability, fewer missed cues, fewer errors at critical moments.
Make transfer the outcome: adapt to teach strategies, not compliance; fade scaffolds when cognition stabilizes (Baceviciute et al., 2022; Doolani et al., 2020).	Prompt dependency and brittle step-following masked by successful completion.	Eye tracking for strategy shifts; combine with physiology when validating retention/transfer costs (Ayres et al., 2021; Souchet et al., 2022).	Fading policies: reduce step prompts as attention stabilizes; add self-check routines; convert hints into reflective prompts under low load. *Primary outcomes to evaluate:* Near/far transfer, retention, robustness to variation, reduced reliance behaviors.

Continued

Table 1: Continued.

Design Guidance	Invisible Cognitive Load Target	Minimal Sensing Needed (Examples)	Adaptive Policy Examples (Teammate-Like Interface)
Design for teaming transparency and trust calibration: brief rationale for adaptations prevents adaptation itself from becoming load (Antonaci et al., 2024; Testa et al., 2025).	Confusion/friction from unexplained system behavior; miscalibrated trust in automation.	Lightweight sensing is sufficient; emphasis is intelligible adaptation, user control, and predictable boundaries (Antonaci et al., 2024).	Micro-explanations ("Reducing overlays to help focus"); user override; predictable adaptation boundaries; mixed-initiative controls. *Primary outcomes to evaluate:* Near/far transfer, retention, robustness to variation, reduced reliance behaviors. Trust calibration, reduced frustration, stable workload/SA under adaptation, user acceptance.

Collectively, these principles reframe "XR + sensing" from a measurement layer into an adaptive teaming capability: a system that regulates information density, pacing, and scaffolding based on sensed cognitive state to protect situation awareness, improve performance robustness, and prioritize transfer. In practical terms, the table emphasizes that effective systems do not require maximal sensor stacks; they require clear cognitive control objectives, parsimonious sensing aligned to those objectives, and transparent adaptation policies that support trust calibration rather than introducing new sources of invisible load (Antonaci et al., 2024; Dey et al., 2019; Testa et al., 2025).

CONCLUSION

XR paired with multimodal sensing is a promising pathway for Industry 5.0 training systems that function as hybrid cognitive-centered systems within complex human–technology ecosystems. Across the included evidence base, XR produced reliable improvements in task performance, while workload effects were mixed and highly design-dependent, reinforcing that "better XR" is not synonymous with "more immersive XR." The core contribution of XR + sensing is its capacity to reveal **invisible cognitive load**: the latent strain, attentional drift, and physiological activation that can accumulate even when trainees appear to perform adequately. Yet the practical goal is not simply to make cognition visible, but to use this visibility to **engineer cognitive readiness** through intelligent interfaces and human-machine teaming.

Several limitations constrain strong conclusions and define the immediate research frontier. First, the evidence base shows operationalization gaps: workload is commonly measured, but situation awareness and transfer remain under-measured and inconsistently defined, limiting synthesis of the outcomes most central to resilient performance. Second, sensing pipelines vary widely across studies in preprocessing, artifact handling, synchronization, and fusion methods, reducing comparability and slowing progress toward reproducible, deployable cognitive-state estimation. Third, truly adaptive systems are scarce; most XR + sensing implementations remain diagnostic-only, meaning the field has not yet accumulated enough closed-loop evaluations to draw firm conclusions about adaptive effectiveness. Finally, ecological trade-offs persist: increasing realism and in-situ deployment enhances industrial relevance but introduces noise and variability in signals, requiring designs and analytic pipelines that tolerate operational conditions rather than assuming laboratory cleanliness.

Future work should therefore shift from instrumentation to integration. The next wave of industrial XR should be designed and evaluated as a coherent human–artificial cognitive system with explicit adaptive policies that translate sensed states into timely, minimal, and transparent changes in guidance, pacing, and information structure. Empirically, this requires (1) stronger and more standardized measurement of situation awareness and transfer alongside workload and performance, (2) clearer reporting and benchmarking of sensing pipelines and fusion decisions, and (3) field-valid trials in situ that test closed-loop adaptation under realistic industrial constraints. If XR + multimodal sensing is to fulfill its human systems integration promise, it will be judged not only by realism or usability, but by whether it reliably regulates invisible cognitive load in real time to protect situation awareness, strengthen performance, and support transfer in the environments where expertise ultimately matters.

REFERENCES

Alessa, F. M., Alhaag, M. H., Al-Harkan, I. M., Nasr, M. M., & Kaid, H. (2023). Evaluating physical stress across task difficulty levels in augmented reality-assisted industrial maintenance. *Applied Sciences, 14*(1), 363.

Antonaci, F. G., Olivetti, E. C., Marcolin, F., Castiblanco Jimenez, I. A., Eynard, B., Vezzetti, E., & Moos, S. (2024). Workplace well-being in Industry 5.0: A worker-centered systematic review. *Sensors, 24*(17), 5473.

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology, 12*, 702538.

Baceviciute, S., Lucas, G., Terkildsen, T., & Makransky, G. (2022). Investigating the redundancy principle in immersive virtual reality environments: An eye-tracking and EEG study. *Journal of Computer Assisted Learning, 38*(1), 120–136.

Beer, P., & Mulder, R. H. (2020). The effects of technological developments on work and their implications for continuous vocational education and training: A systematic review. *Frontiers in Psychology, 11*, 918.

Biondi, F. N., Cacanindin, A., Douglas, C., & Cort, J. (2021). Overloaded and at work: Investigating the effect of cognitive workload on assembly task performance. *Human Factors, 63*(7), 813–820.

Debie, E., Fernandez Rojas, R., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., & Abbass, H. A. (2021). Multimodal fusion for objective assessment of cognitive workload: A review. *IEEE Transactions on Cybernetics, 51*(3), 1542–1555.

Dey, A., Chatburn, A., & Billinghurst, M. (2019). Exploration of an EEG-based cognitively adaptive training system in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 220–226). IEEE.

Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., & Makedon, F. (2020). A review of extended reality (XR) technologies for manufacturing training. *Technologies, 8*(4), 77.

Fazli, B., Sajadi, S. S., Jafari, A. H., Garosi, E., Hosseinzadeh, S., & Zakerian, S. A. (2025). EEG-based evaluation of mental workload in a simulated industrial human–robot interaction task. *Health Scope, 14*(1).

Gavish, N., Gutierrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments, 23*(6), 778–798.

Grassini, S., & Laumann, K. (2020). Evaluating the use of virtual reality in work safety: A literature review. In P. Baraldi, F. Di Maio, & E. Zio (Eds.), *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference* (pp. 4964–4971). Research Publishing.

Gualtieri, L., Fraboni, F., De Marchi, M., & Rauch, E. (2022). Development and evaluation of design guidelines for cognitive ergonomics in human–robot collaborative assembly systems. *Applied Ergonomics, 104*, 103807.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 52, pp. 139–183). Elsevier.

Howard, M. C., & Davis, M. M. (2023). A meta-analysis of augmented reality programs for education and training. *Virtual Reality, 27*(4), 2871–2894.

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior, 124*, 106808.

Limbu, B. H., Jarodzka, H., Klemke, R., & Specht, M. (2018). Using sensors and augmented reality to train apprentices with recorded expert performance: A systematic literature review. *Educational Research Review, 25*, 1–22.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236.

Mondellini, M., Pirovano, I., Colombo, V., Arlati, S., Sacco, M., & Mastropietro, A. (2024). A multimodal approach exploiting EEG to investigate the effects of virtual reality environment on mental workload. *International Journal of Human–Computer Interaction, 40*(7), 6566–6578.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71.

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest, 13*(2), 74–101.

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human–Computer Interaction, 38*(8), 801–824.

Testa, A., Vagnetti, A., Paoli, A., Agostini, M., Gal, B., & Pausder, H. (2025). Fuzzy-based sensor fusion for cognitive load assessment in inclusive manufacturing strategies. *Sensors, 25*(11), 3356.

Vanneste, P., Raes, A., Bombeke, K., Van Acker, B. B., & Van den Noortgate, W. (2021). Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work, 23*(2), 567–585.

Zakeri, Z., Arif, A., Omurtag, A., Breedon, P., & Khalid, A. (2023). Multimodal assessment of cognitive workload using neural, subjective and behavioural measures in smart factory settings. *Sensors, 23*(22), 8926.