RESEARCH / ARTUN SEPKEN

Designing for Interrupted Attention

Adaptive interaction design under uncertain evidence

Artun Sepken

Research dossier | July 2026

Independent student research developed at FH JOANNEUM. The prototype and pilot are exploratory; they do not establish accurate attention detection or general performance effects.

Contents

Practitioner Summary

Digital systems increasingly decide when to request attention, when to offer assistance and when to act without waiting. This dossier asks a design question that sits underneath those behaviours: how should an adaptive interface respond when its evidence about a person's capacity is incomplete?

The research began with flow and interruption, then expanded through cognitive load, emotion, memory, workload measurement, physiological computing and mixed-initiative systems. Fifteen research posts, a large Zotero library, a public presentation and two prototypes were used to turn that field into a more specific position.

Four conclusions guide the work:

- Interruption cost includes recovery, not only the moment of distraction.
- Measures such as EEG, behaviour and task demand are proxies, not access to a

user's mental state.

- The authority of an adaptation should be proportional to the certainty and

consequence of the evidence behind it.

- Useful adaptation should remain legible, negotiable, reversible and open to

meaningful override.

The neuroadaptive UAV console described later is one research probe inside this larger programme. Its small pilot produced mixed observations: lower reaction times in the working-build adaptive condition, but lower accuracy, longer resumption lag and no clear workload benefit. The value of the probe is not a claim that adaptation worked. It is the way the prototype exposed the design consequences of uncertain sensing.

1. Research Trajectory

Beginning with flow, not distraction

Additional synthesis: how the research question changed

Phase	Focus	What it changed
Flow and continuity	Conditions for sustained, meaningful interaction	Continuity, feedback and felt control became design concerns; a deficit model of the user was rejected
Interrupted activity	Notifications inside fragmented knowledge work	The unit of analysis moved from the alert to the activity around it
Cognitive and emotional cost	Workload mechanisms and affective response	Interruption cost widened beyond completion time to frustration, trust and perceived respect
Memory and recovery	Goal decay and task resumption	Recovery reframed as shared work between person and interface
The adaptive turn	Closed-loop and mixed-initiative systems	The question became how systems should act on uncertain evidence about a changing person
Public explanation and making	UX Graz talk and neuroadaptive probe	A broad field map narrowed to a testable interaction logic

Source: Consolidated from the DR2 blog sequence and research archive.

The earliest work was organised around a positive question: what conditions allow sustained, meaningful interaction? Flow theory supplied an initial vocabulary of clear goals, feedback, challenge-skill balance, concentration, and control. This mattered because it prevented the inquiry from beginning with a deficit model of the user. The problem was not assumed to be that people had become incapable of focusing.

The problem was that interactive systems participate in creating or breaking the conditions under which focus is possible.

The first two posts revisit Csikszentmihalyi through an interaction-design lens. They also expose an early ambiguity. "Flow" can refer to a demanding psychological state, but in design practice it is often reduced to frictionlessness or smooth task completion. The research gradually moved away from treating flow as a state that an interface should manufacture. Its lasting value was more modest: flow made continuity, feedback, and felt control visible as design concerns.

This distinction remains important. An interface cannot promise optimal experience, and constant engagement is not evidence of flow. Yet interfaces do structure goals, tempo, feedback, and transitions. They can make concentration easier to maintain or repeatedly force users to reconstruct what they were doing. Flow was therefore not the thesis answer. It was the conceptual starting point from which interruption became meaningful.

From notification events to interrupted activity

The notification and fragmented-work posts shifted the unit of analysis. Notifications are visible examples of interruption, but the meaningful object is the activity in which they arrive. Research by Mark, Gonzalez, and Harris (2005) and Iqbal and Horvitz (2007) made clear that knowledge work already consists of interleaved activities, self-interruptions, and chains of diversion. A message may take seconds to read while producing a much longer reorientation process.

The notification research also introduced a social tension. Turning notifications off can support productivity, yet it can produce anxiety or feelings of disconnection (Pielot & Rello, 2017). Deferral and batching can reduce disruption, but no fixed rule resolves every situation. A family message, a safety alert, a team request, and a promotional prompt impose different obligations even when they share the same visual component. The research could no longer treat interruption as a quantity to minimise. It had to consider urgency, source, relationship, task phase, and consequence.

The taxonomy post was an early attempt to organise this complexity. Interruption varies by modality, initiator, timing, coordination method, and environment. McFarlane's (2002) comparison of immediate, negotiated, mediated, and scheduled interruption strategies later gave this intuition a more precise interaction vocabulary. Taxonomy became useful not as classification for its own sake, but as protection against designing one universal solution for events that differ structurally.

Cognitive and emotional cost

The cognitive-load post complicated the simple idea that busy users should receive fewer alerts. Lavie's (2010) load theory distinguishes perceptual demand from demands on working memory and cognitive control. High perceptual load may reduce processing of irrelevant stimuli, while high working-memory load can make distraction harder to resist. Workload is not a single visible quantity, and a system that labels somebody "busy" may collapse several different mechanisms.

The emotional-design post widened the cost model further. Interruptions communicate more than content. Sound, vibration, placement, repetition, and timing can signal urgency, authority, intimacy, or obligation. Norman's visceral, behavioural, and reflective levels helped describe why the same message can feel supportive in one moment and hostile in another. Affect also accumulates. A system may never produce one catastrophic interruption while still teaching the user that they are permanently reachable and never fully in control.

This was a decisive shift in the research. Interruption could no longer be evaluated only through completion time. Frustration, pressure, trust, perceived respect, and social expectation became part of the interaction. Adamczyk and Bailey (2004) are especially valuable because they connect timing with both performance and affective response. Their participants did not merely work differently at different interruption moments; they also judged the interrupting system differently.

Memory, recovery, and the return

The memory and recovery work moved attention from the interruption to its aftermath. Altmann and Trafton's (2002) memory-for-goals model describes suspended goals as activation-dependent structures rather than intact objects waiting to be reopened. Borst, Taatgen, and van Rijn (2015) add a process account in which resumption is disrupted when the problem state cannot be maintained or easily reconstructed. These models gave design weight to a familiar experience: returning to the same screen is not the same as returning to the same thought.

This insight produced one of the most durable positions in the archive: memory is partly distributed across person and interface. Stable state, visible history, progress markers, annotations, and cues can reactivate the goal structure that an interruption displaced. Chen et al. (2025) further suggest that cues oriented toward previous state and cues oriented toward upcoming action can shape recovery differently. The appropriate cue depends on what the user has lost: location, intention, reasoning, or next step.

Recovery later became one candidate thesis direction, especially for AI-assisted work, but the June 2026 research-state note explicitly corrects an over-narrow interpretation. Cognitive re-entry is one strong thread, not the declared identity of the whole project. The umbrella remains adaptive interaction design.

The adaptive turn

The EEG workshop and neuroadaptive post changed the form of the question. Closed-loop physiological systems suggested that interfaces might respond to changes in human capacity rather than deliver the same behaviour at every moment. The attraction was not a belief that consumer EEG could read attention. It was the interaction logic: sense, infer, adapt, observe again.

Historical work made clear that this logic did not begin with contemporary AI. Pope, Bogart, and Bartolome (1995) used an EEG-derived engagement index to alter automation in a flight-deck task. Fairclough (2009) formalised physiological computing as a loop in which an inferred user state becomes input to system adaptation. Horvitz (1999) framed mixed initiative as a decision about when a system should act, ask, suggest, or remain quiet. These projects differ in technology and context, but they share a concern with distributing initiative over time.

The research direction was consequently renamed adaptive interaction design. This label is intentionally broader than neuroadaptive interface design. Adaptation may use task state, interaction behaviour, context, user preference, AI inference, or physiology. The central issue is not which sensor appears most futuristic. It is how a system should behave when it has incomplete evidence about a changing person and situation.

Research through public explanation and making

The UX Graz invitation and the prototype transformed the archive from private reading into accountable practice. The talk demanded a coherent public argument. The prototype demanded operational choices: which signals count, how they are combined, what thresholds trigger change, what the system does, and what control remains with the user.

Both activities narrowed the research, but in different ways. The talk revealed that a broad field map needs a concrete object. The prototype revealed that an apparently concrete object still contains many theoretical decisions. Together they produced the present position: adaptive systems should not only optimise intervention timing; they should make intervention legible, negotiable, reversible, and open to meaningful override.

How the literature changed the research question

Seven clusters from the local evidence library, each adding a different design obligation.

RESEARCH CLUSTER	ANCHOR SOURCES	QUESTION INTRODUCED
Flow and attention	Csikszentmihalyi (1988); Lavie (2010)	What conditions support sustained activity, and when does demand protect or disrupt focus?
Interruption timing	Adamczyk & Bailey (2004); McFarlane (2002)	When should an interruption occur, and who should coordinate its timing?
Memory and resumption	Altmann & Trafton (2002); Borst et al. (2015); Chen et al. (2025)	What context must remain available so a person can reconstruct the suspended goal?
Workload measurement	Kosch et al. (2023); Koundal et al. (2024)	Which measurements are suitable, and what can they legitimately support?
Physiological computing	Fairclough (2009); Pope et al. (1995)	How should uncertainty in sensed signals constrain system authority?
Mixed initiative	Horvitz (1999); Shneiderman & Maes (1997)	How should initiative move between person and system without erasing agency?
Context and embodiment	Dourish (2004); Liao et al. (2020)	How can adaptation remain situated, intelligible and open to challenge?

Source: Author evidence map from the annotated DR2 library and fifteen-post research sequence.

Figure 2. The evidence map connects each literature cluster to anchor sources and the design question it introduced.

Figure source: Author evidence map from the local Zotero library and fifteen-post research sequence.

2. Literature Landscape

The 132-record library is organised here as a landscape of connected problems, not as a flat bibliography. Seven clusters have been most consequential.

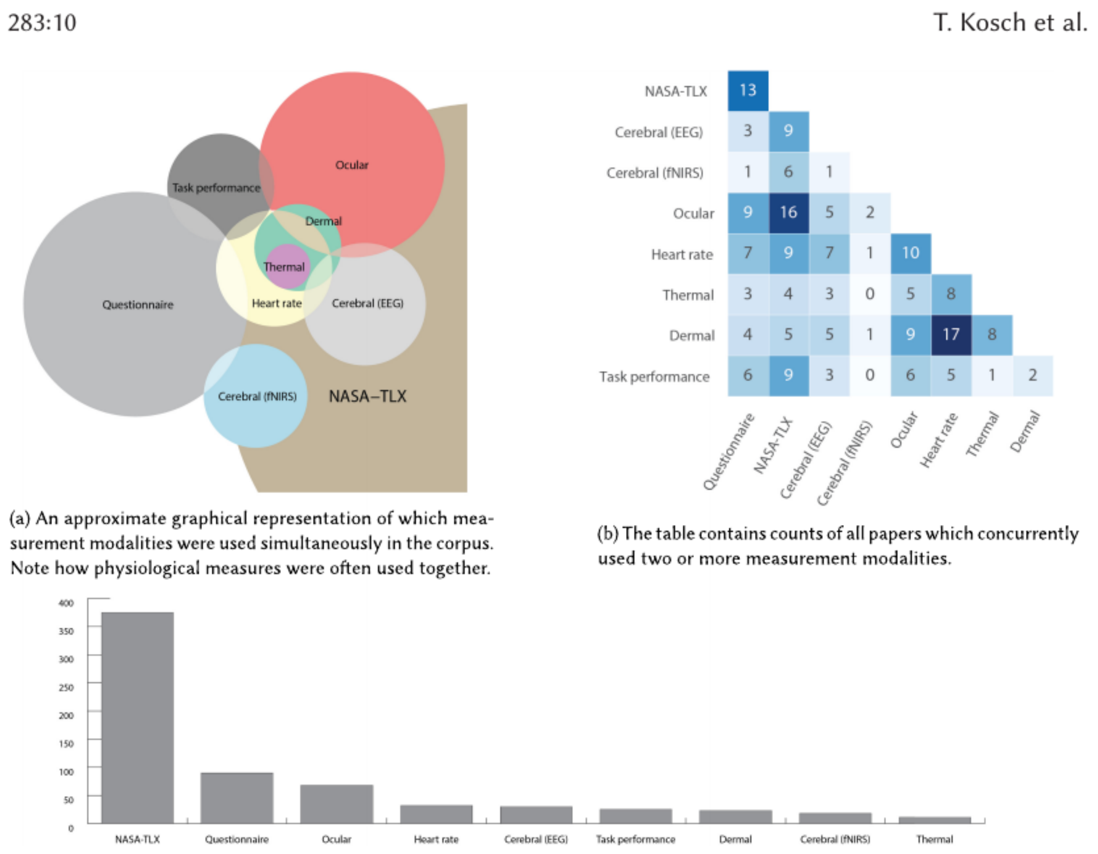
Attention, flow, and temporal experience. Csikszentmihalyi supplies the positive vocabulary of absorbed, goal-directed activity. Bradbury (2016) is important for rejecting simplistic claims that human attention has one short universal duration. Liikkanen and Gomez (2013) extend the issue into interaction design by asking how systems shape the experience of time. Together, these sources support a careful premise: attention is contextual, and interface behaviour contributes to its continuity without fully determining it.

Interruption timing and coordination. Adamczyk and Bailey (2004), McFarlane (2002), and Iqbal and Bailey (2005) examine when and how interruption should be coordinated. The literature repeatedly shows that timing is not neutral. Task boundaries, workload, interruption method, and urgency change both cost and acceptability. Fogarty and colleagues' sensor-based interruptibility work demonstrates that context inference is possible to a degree, while also exposing the gap between statistical prediction and a socially appropriate intervention.

Fragmented work, memory, and resumption. Mark et al. (2005) describe fragmented knowledge work; Iqbal and Horvitz (2007) follow disruption and recovery in everyday computing; Altmann and Trafton (2002) explain the retrieval dynamics of suspended goals; Borst et al. (2015) model the problem-state bottleneck; and Chen et al. (2025) examine recovery cues. This cluster changes the object from a message to a trajectory: preparation, diversion, memory decay, reorientation, and continuation.

Notification systems and digital attention. Pielot and Rello (2017), Weber et al. (2018), Fitz et al. (2019), Mehrotra and Musolesi (2018), Okoshi et al. (2015), and Li et al. (2022) explore suppression, deferral, batching, handling behaviour, and preference prediction. Monge Roffarello and colleagues add an ethical design layer by identifying attention-capture patterns and proposing digital-attention heuristics. This body of work shows both the value and the limitation of notification-centred interventions. Delivery can be improved, but sender priorities, platform incentives, and social expectations still shape the interruption.

Cognitive workload and physiological computing. Sweller and Chandler (1991) provide a cognitive-load foundation; Lavie (2010) distinguishes mechanisms of load and distraction; Koundal et al. (2024) review interruption and mental workload; Kosch et al. (2023) survey workload measurement in HCI. Pope et al. (1995), Fairclough (2009), Frey et al. (2013), Afergan et al. (2014), and Treacy Solovey et al. (2015) examine physiological input and adaptive response. The key lesson is not that one signal reveals the mind. It is that multiple imperfect measures can inform modest adaptation if uncertainty and individual difference are taken seriously.



	Questionnaire	NASA-TLX	Cerebral (EEG)	Cerebral (fNIRS)	Ocular	Heart rate	Thermal	Dermal
NASA-TLX	13							
Cerebral (EEG)	3	9						
Cerebral (fNIRS)	1	6	1					
Ocular	9	16	5	2				
Heart rate	7	9	7	1	10			
Thermal	3	4	3	0	5	8		
Dermal	4	5	5	1	9	17	8	
Task performance	6	9	3	0	6	5	1	2

(a) An approximate graphical representation of which measurement modalities were used simultaneously in the corpus. Note how physiological measures were often used together.

(b) The table contains counts of all papers which concurrently used two or more measurement modalities.

(c) The graph visualizes the total number of times each modality was employed in the corpus. Questionnaire-based methods were most prevalent, with a big emphasis on using the NASA-TLX.

Fig. 2. An overview of the different modalities used to estimate cognitive load in the corpus.

Figure 3. Kosch et al.'s survey shows the range and overlap of modalities used to estimate cognitive workload in HCI.

Figure source: Reproduced from Kosch et al. (2023), Figure 2, "A Survey on Measuring Cognitive Workload in Human-Computer Interaction", ACM Computing Surveys 55(13s), DOI 10.1145/3582272.

Mixed initiative, proactive systems, and agentic AI. Horvitz (1999) and the Shneiderman-Maes debate establish an older argument about direct manipulation, agents, and shared initiative. Contemporary work on proactive assistants and multi-step agents reopens this problem at a larger scale. Zhou et al. (2026) ask when users should check agentic work; Epperson et al. (2025) examine debugging and steering; Mozannar et al. (2025) design human-in-the-loop agentic interaction; and Zindulka et al. (2026) address memory errors around AI-supported creation. Here, interruption is not always a message. It may be a confirmation request, an autonomous action, a handoff, or the need to reconstruct what an agent changed.

Context, spatial computing, and embodied environments. Dourish (2004) warns that context is not merely a stable set of variables waiting to be captured. Weiser, Dourish, Suchman, Scaife and Rogers, and Ishii and Ullmer provide broader accounts of situated, embodied, and externalised interaction. More recent mixed-reality studies explore notification timing, placement, distraction reduction, and situation awareness in head-mounted displays. This cluster matters because always-on and spatial interfaces can place system demands directly into the perceptual field. It remains an application context rather than the fixed thesis identity.

Across these clusters, the library changes the research in two ways. First, it prevents adaptive interaction from being treated as a novel reaction to generative AI. The underlying problems of context, initiative, workload, interruption, and human control have a long HCI history. Second, it prevents one literature from dominating. Notification research alone cannot answer how autonomous agents should act. Physiological computing alone cannot determine what a respectful intervention looks like. Memory research alone cannot settle questions of authority. The research question sits at their intersection.

3. What the Literature Changed

The literature changed the project in three stages. First, it moved the inquiry from interruption frequency to interruption timing and recovery. Second, it replaced the idea of a knowable user state with a measurement problem. Third, it reframed adaptation as a negotiation of initiative rather than a technical response to a sensor value.

From delivery to continuity

Interruption research initially suggested a straightforward design response: avoid badly timed prompts. Adamczyk and Bailey (2004) support the importance of task boundaries, while Iqbal and Horvitz (2007) show the long tail that can follow disruption in everyday work. Together, these studies make a stronger argument than "notifications are annoying." The cost of an interruption is distributed across emotional response, task switching, memory, and recovery.

Altmann and Trafton (2002) sharpened the design interpretation. If resumption depends on retrieving a decaying goal, preventing an interruption is only one intervention point. Interfaces can also support prospective memory before a switch, preserve task state during it, and provide cues after it. This turns recovery from an individual coping skill into a shared responsibility between person and system.

More recent work on recovery cues adds specificity. In a concurrent multitasking environment, Chen et al. (2025) compared cues that pointed back to prior state with assistant cues oriented toward the next action. Their findings do not establish a universal cue pattern, but they show that cue type matters and that performance, workload, and alertness may not move together. The design implication is not "always tell the user what to do next." It is that recovery support should be selected according to the task, the consequence of error, and the kind of continuity that has been lost.

This body of work also resists a binary view of interruptions as good or bad. Some events are urgent; some preserve social connection; some prevent costly failure. A system that suppresses every demand for attention would be as crude as one that delivers all demands immediately. The design space includes immediate interruption, negotiated handling, system-mediated delay, batching, peripheral presentation, and post-interruption support. The appropriate form depends on urgency, consequence, task phase, modality, and the user's ability to recover.

From "user state" to uncertain evidence

The second change came from workload research. Cognitive workload is relevant because an adaptive system needs some basis for deciding whether to remain quiet, offer help, or change automation. Yet workload is not equivalent to task difficulty, stress, effort, fatigue, or attention. These constructs overlap without becoming interchangeable. Kosch et al. (2023) show that HCI research measures workload through combinations of self-report, performance, behaviour, and physiology. Each family of measures captures a different part of the situation and introduces its own limitations.

This makes a single "attention score" epistemically and operationally dangerous. A low response rate may indicate overload, confusion, disengagement, or a deliberate choice. A physiological change can have several causes, while apparently similar experiences can produce different signals across people. Retrospective self-report can reveal felt workload but is not a continuous, objective ground truth. Task-demand models can describe the designed environment but not fully explain how an individual experiences it.

Fairclough's (2009) model of physiological computing made the uncertainty structurally visible. A biocybernetic loop senses data, infers a state, adapts the system, and observes the resulting response. The inference stage is not a transparent bridge from body to mind. It is a model-dependent interpretation. Once the system acts on that interpretation, it also changes the conditions it is trying to measure. Adaptation is a feedback loop, not a one-time classification.

Treacy Solovey et al. (2015) translate this limitation into interaction guidance. Physiological signals are often better suited to implicit context than to direct commands; delays and uncertainty favor gentle adaptations, confidence thresholds, and reversibility. The important lesson is proportionality. An uncertain input should not produce an irreversible or high-authority response. Confidence is not only a number for the sensing pipeline. It should shape what the interface is permitted to do.

From automation to negotiated initiative

The third change concerned agency. Mixed-initiative interaction predates current AI systems. Horvitz (1999) describes systems that infer context, offer actions, and share initiative with users. The persistence of the problem is instructive: technical capability does not determine the right moment or level of intervention. A system may be able to act and still have insufficient grounds to do so.

This becomes more difficult when the intervention is meant to protect attention. Constant confirmation can itself become an interruption. Silent automation can preserve short-term focus while obscuring what changed. The choice is not simply user control versus system control. It is a continuing allocation of initiative, with different requirements for visibility, consent, correction, and re-entry.

The literature therefore produced a set of design criteria rather than a sensing target:

- **Legibility:** the user can see that adaptation occurred and understand its practical reason.
- **Negotiability:** the user can accept, defer, question, or reshape the system's intervention.
- **Reversibility:** the effects of adaptation can be undone without disproportionate recovery work.
- **Override:** manual authority remains available when the estimate or intervention feels wrong.
- **Continuity:** the system preserves enough state and provenance for the user to return.

These criteria are a design interpretation of the literature, not a validated universal checklist. Their value is that they redirect evaluation. The question becomes not only whether an estimator classified a state correctly, but whether the interaction remains accountable when classification is uncertain.

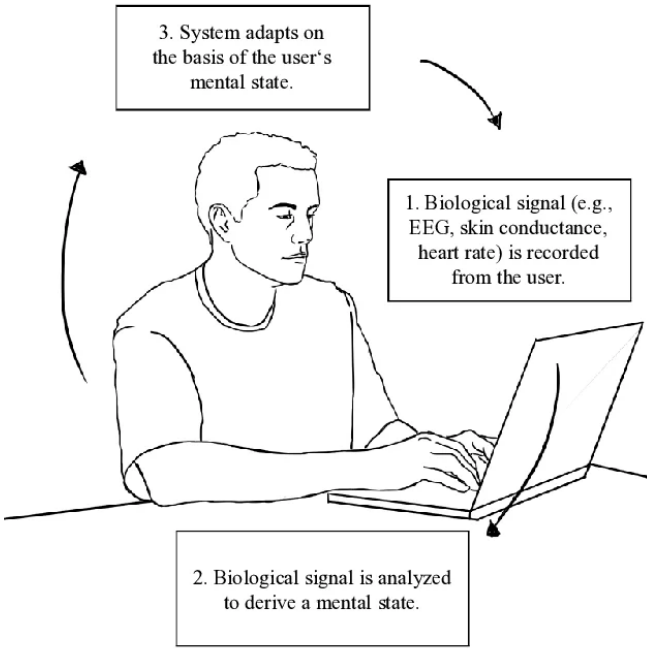


Figure 4. Riedl and Leger's conceptual illustration makes the neuroadaptive loop explicit: record a biological signal, infer a state, and adapt the system.

Figure source: Reproduced from Riedl and Leger (2016), "Fundamentals of NeuroIS", p. 18; archived in the DR2 research workspace.

4. Prototype as One Research Probe

The resulting probe was a neuroadaptive ground-control interface in which an operator supervises semi-autonomous unmanned aerial vehicles. The scenario was selected because it concentrates competing priorities, timed contingencies, and changing task demand in one understandable environment. It was not selected because drone supervision defines the thesis. The operational setting is a container for studying adaptive behaviour under visible pressure.

Additional evidence boundary: what each prototype input measures

Input	What it measures	What it is not
Muse EEG	An engagement index derived from a real EEG stream	Ground truth about attention or mental state
Webcam proxy	Face presence, blink rate and head motion	Calibrated point-of-gaze tracking
Task demand	Mission state, unassigned UAVs and unresolved-alert backlog	The operator's subjective experience
Signals are normalised against a short per-participant baseline and fused by a rule-based estimator into three interface states: underload, stable and overload.		

Source: Prototype records and implementation notes.

The operator's task was to keep the UAVs functioning while assigning work and responding to contingencies such as link loss, low fuel, traffic, or geofence events. Contingencies arrived in phases designed to create transient demand. This made interruption timing consequential: some events required action, while lower-urgency information could potentially be deferred.

The prototype estimated underload, stable operation, and overload from three imperfect inputs. A Muse headband supplied real EEG data. A webcam-derived measure combined face presence, blink rate, and head motion as an attention proxy; it was not calibrated point-of-gaze. A task-demand component represented mission state, including UAVs awaiting tasking and unresolved alerts. Signals were normalised against a short per-participant baseline and combined in a rule-based estimator.

This pipeline must be described precisely. The system did not detect attention accurately, know the operator's mental state, or treat EEG and webcam data as ground truth. It produced an estimate from proxies. The categories were interface states generated by a model, not diagnoses of the participant.

In the adaptive condition, the system could defer low-urgency interruptions and automatically task at-risk UAVs when estimated load was high. In the static condition, it did not apply those changes. The adaptive behaviour was designed around control negotiation. Operators could inspect the estimated state and retain manual authority. In conceptual terms, the probe asked whether an interface could change its level of assistance while making that change visible and challengeable.

The neuroadaptive form brought the literature into one loop. Fairclough's (2009) sequence of sensing, inference, adaptation, and feedback became an implemented interaction. Pope, Bogart, and Bartolome's (1995) earlier use of an EEG-derived engagement index in an automated task provided historical precedent, while also underscoring that specialised measurements should not be transferred uncritically into general interfaces. Treacy Solovey et al. (2015) supported treating physiology as uncertain context rather than a direct command channel.

Building the probe also revealed that sensor realism does not guarantee research validity or interaction quality. Early sessions showed that the estimator rarely entered overload, so the adaptive layer had little opportunity to act. The system was then recalibrated, a task-backlog term and per-participant baseline were added, and logging was improved. These changes made the adaptive mechanism more active, but they also divided the study into different software states. The first sessions became evidence about prototype iteration; they could not be pooled uncritically with later sessions as if all participants had experienced the same system.

This is why "prototype as research probe" is the appropriate frame. The artifact operationalized assumptions and made their consequences inspectable. It demonstrated an end-to-end connection among real sensors, task context, an estimator, and adaptive interface behaviour. It also exposed fragility across participants, dependence on calibration, and tension between faster handling and correct handling. Its value lies in what it made available for critique, not in establishing a finished adaptive solution.

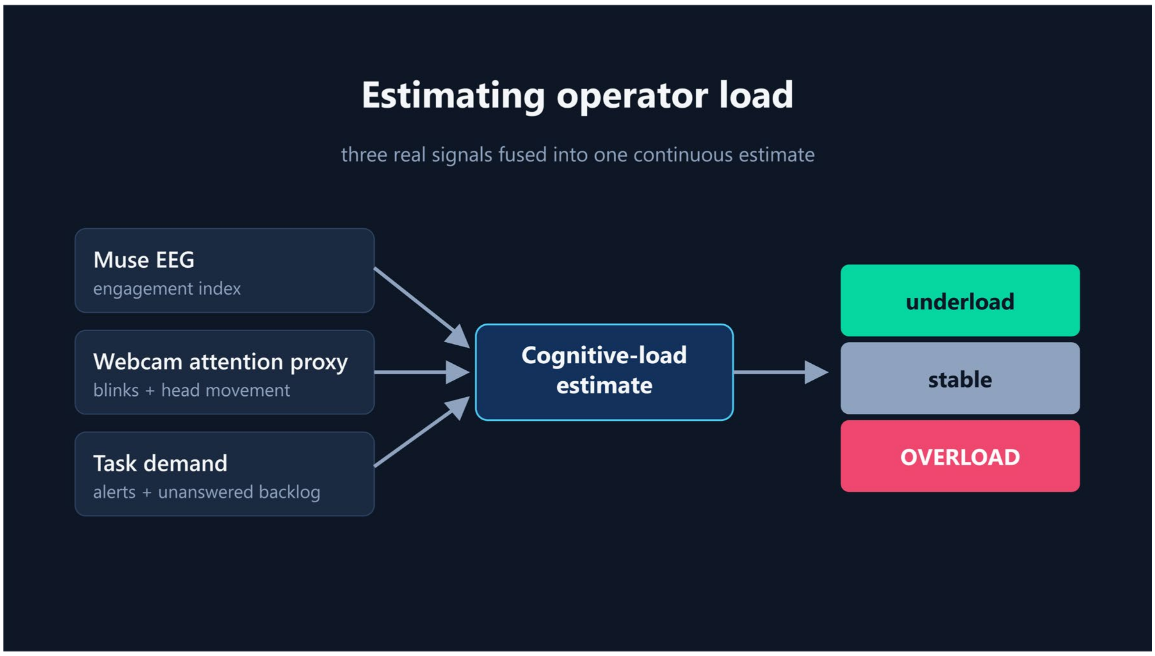


Figure 5. Prototype signal fusion combined Muse EEG, a webcam attention proxy and designed task demand.

Figure source: Author diagram based on the implemented prototype architecture.

PROTOTYPE PROBE

What changed between the two experimental conditions

RESEARCH VARIABLE	ADAPTIVE CONDITION	STATIC CONDITION
Task environment	Same supervisory task	Same supervisory task
Workload estimate	Used by adaptation logic	Recorded, no interface effect
Low-priority alerts	Could be deferred	Delivered immediately
Assistance	Could be offered	No adaptive assistance
Operator authority	Inspect and override	Direct control

Figure 6. The prototype probe held the task environment constant while changing how workload estimates affected alerts and assistance.

Figure source: Author condition matrix derived from the implemented prototype, June 2026.

5. What the Pilot Can and Cannot Claim

Four participants completed both adaptive and static conditions in a small within-subjects Wizard-of-Oz pilot. The prototype changed between sessions. The first two sessions used formative builds in which overload was rarely reached and adaptation did not reliably engage. Both formative participants completed the adaptive condition first, creating an order confound; counterbalancing was applied only to the later working-build pair. With N = 4, differing builds, this order limitation, and coarse measures, the study is small, formative, and descriptive. It does not support statistical inference.

Additional pilot summary: direction of the working-build observations

Measure	Direction on the working build (adaptive vs static)	Interpretation boundary
Interruption reaction time	Shorter in the adaptive condition	Consistent with the intended deferral mechanism; two participants only
Response accuracy	Lower in the adaptive condition	Indicates a speed–accuracy trade-off
Resumption lag	Longer in the adaptive condition	Opposite of the intended effect
Self-reported workload (NASA-TLX)	No clear difference	No subjective benefit demonstrated

Source: Author analysis of the four-participant pilot; descriptive only.

The pilot can claim that an interface using real Muse EEG, a webcam attention proxy, and task-demand data was built and operated end to end. It can describe what happened in the recorded sessions. On the working build, low-urgency deferral in the adaptive condition coincided with shorter mean interruption reaction times than in the static condition. This observation is consistent with the intended mechanism: spreading lower-urgency alerts can reduce pile-up during demanding periods.

That observation cannot be generalized as proof that adaptation improves performance. Accuracy was lower in the adaptive condition on the working build, indicating a speed-accuracy trade-off. Resumption lag moved in the opposite direction from the hoped-for effect. Self-reported workload did not show a clear adaptive benefit. With only two participants using the working build, these differences are prompts for investigation, not stable effect estimates.

The pilot also cannot claim that the system accurately detected attention, cognitive workload, or any participant's internal state. The webcam measure was an attention proxy based on face presence, blink rate, and head movement, not gaze direction. The EEG measure was transformed through a chosen engagement index. Task demand represented conditions designed into the mission, not subjective experience. The estimator fused these quantities with fixed weights and thresholds. Its output was a design instrument, not ground truth.

Individual traces illustrate why this distinction matters. Fixed rules could under-trigger for one participant and remain near overload for another. Per-participant normalisation helped make adaptation operational, but it did not remove ambiguity. A system that is almost always "overloaded" loses the phasic contrast needed to time interventions; one that never reaches overload cannot adapt at all. These are not minor calibration inconveniences. They are evidence that the interaction must remain usable when the model is uncertain or wrong.

The strongest claim is therefore methodological and design-oriented. The pilot made it possible to observe an adaptive loop, identify where the mechanism failed to engage, and compare intended support with unintended costs. It showed why evaluation must examine more than one measured outcome. Reaction speed, response accuracy, resumption, and subjective workload describe different aspects of the experience, and improving one does not establish overall benefit. Questions of legibility, understanding, trust, and perceived control remain priorities for a future qualitative evaluation; they were not measured in this pilot.

For the next study, the build should be frozen before recruitment, condition order fully counterbalanced, the sample expanded, and measures refined. Calibration should update over time rather than depend only on a brief resting baseline. The design should separate urgency, response speed, correctness, and recovery more carefully. Qualitative inquiry should also examine whether participants noticed adaptation, understood its reason, trusted it appropriately, and felt able to override it. These questions are essential because an adaptation that works invisibly according to a metric may still undermine agency.

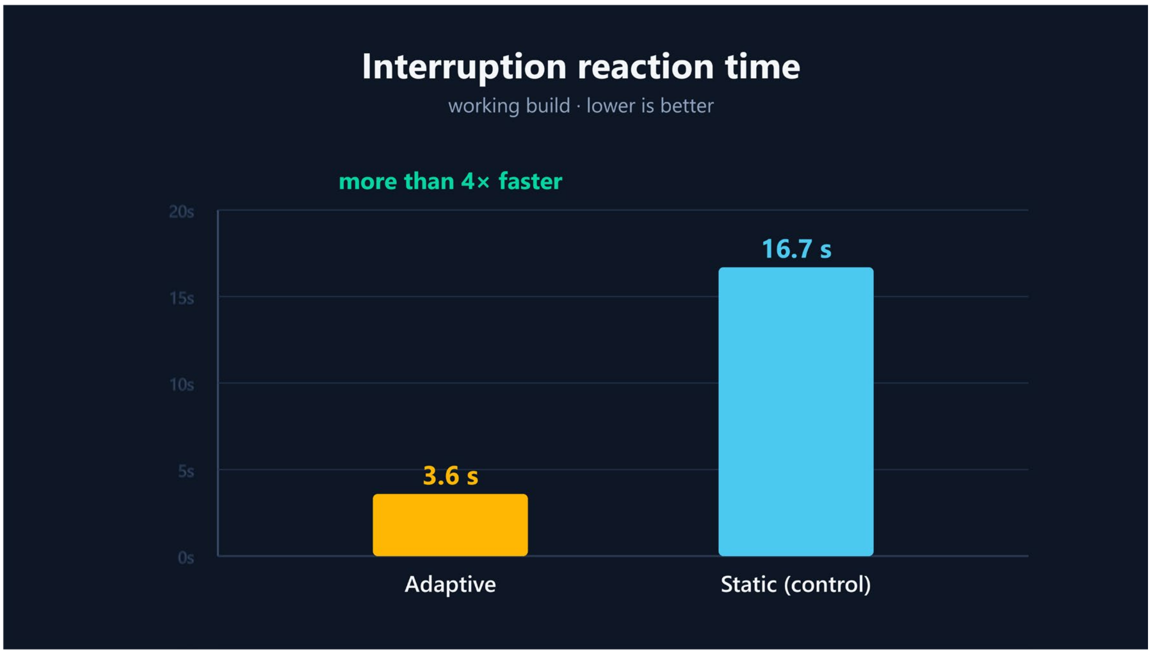


Figure 7. Working-build reaction time is shown descriptively for two participants; it is not an inferential effect estimate.

Figure source: Author analysis of P03-S and P04-EM pilot exports.

6. Design Position

The continuing thesis direction is adaptive interaction design, not neuroadaptive UAV control. The prototype is one research object within that direction. It made uncertain sensing concrete, but the underlying problem also applies to AI agents, collaborative tools, mixed-reality systems, and operational interfaces that infer when to ask, wait, assist, or act.

The next question is narrower than "How can interfaces adapt to attention?" and more accountable than "Can workload be detected?" A productive formulation is:

How can adaptive interfaces respond to uncertain estimates of human capacity while making intervention legible, negotiable, reversible, and open to meaningful override?

This question places uncertainty inside the interaction rather than treating it only as a model-performance issue. A confidence value hidden in a technical pipeline does little for the person affected by the adaptation. Uncertainty becomes a design concern when it changes the intervention's scale, timing, explanation, and reversibility. Low-confidence estimates may justify peripheral cues or offers of assistance, but not silent, consequential action. Higher-confidence estimates still require limits where errors would be costly.

Legibility does not require exposing raw physiological streams or algorithmic detail. It requires a usable account of what changed and what that change means. A message such as "low-priority alerts delayed during a high-demand period" may be more useful than an unexplained workload score. Provenance and history can help users reconstruct system behaviour after the fact, especially when adaptation occurred while their attention was elsewhere.

Negotiability extends beyond a binary approval dialog. Requiring confirmation for every adjustment would reproduce the interruption problem. Negotiation can include setting thresholds, choosing categories that may be deferred, temporarily pausing adaptation, correcting a mistaken estimate, or indicating a preferred level of assistance. The timing and form of that participation should match the consequence of the system's action.

Reversibility concerns both technical state and cognitive recovery. An undo function may restore a file without restoring the user's understanding of why it changed. Reversible adaptation should preserve prior state, reveal the path of intervention, and make return comprehensible. Override similarly means more than a prominent button. Manual control must remain practically usable, and exercising it should not impose a punitive amount of reconstruction.

Human agency ties these qualities together. Agency is not preserved merely because the person remains nominally "in the loop." If the system acts too quickly to inspect, provides no intelligible reason, or makes correction prohibitively difficult, formal control has little substance. Conversely, preserving agency does not require refusing automation. It requires designing automation as a participant in a relationship whose initiative can be examined and revised.

Future work should compare interaction patterns under controlled uncertainty rather than pursue one supposedly definitive attention classifier. Studies could vary the visibility of confidence, the reversibility of intervention, or the point at which users can negotiate adaptation. Behavioural measures should be paired with participant accounts of understanding, control, and recovery. Different contexts should be selected according to consequence: a writing tool may tolerate gentle speculative assistance, while a safety-critical dashboard demands stronger provenance and more conservative authority.

The research began with interrupted attention, but its larger concern is the quality of human-system coordination. As systems become more proactive, they will increasingly manage not only information but also timing, initiative, and the conditions under which people can act. Designing responsibly for that future means neither pretending that uncertainty can be eliminated nor placing all recovery work on the user. It means building systems that can offer support without claiming privileged access to the mind, and that can adapt without making human control ceremonial.

7. Research Agenda

The next phase should narrow the project without discarding the breadth that made the question possible. Five connected studies could turn the current position into a defensible thesis programme.

Study 1: Understand how people interpret adaptation

Before testing another estimator, a qualitative study should examine how users notice and interpret adaptive behaviour. Participants could experience several small interventions in a familiar work tool: delayed low-priority messages, reduced information density, a suggested pause, or a short return cue. Interviews and stimulated-recall sessions would ask what participants believed the system noticed, why they thought it acted, and whether they felt able to challenge it.

The purpose would not be to discover whether people "like adaptation." It would identify the explanations, controls, and mismatches that make adaptation feel helpful, paternalistic, confusing, or invisible. This study would directly test legibility and negotiability.

Study 2: Compare intervention authority under uncertainty

A controlled prototype could vary two factors: confidence in the system's estimate and consequence of the proposed action. Low-authority interventions might include peripheral cues or offers of assistance. Higher-authority interventions might delay information, alter automation, or act on the user's behalf. Participants would compare interfaces that expose uncertainty in different ways.

The central question would be proportionality: how should uncertain evidence constrain what the system is allowed to do? Outcomes should include comprehension, appropriate trust, correction behaviour, task performance, and perceived control, rather than classifier accuracy alone.

Study 3: Design for re-entry after system action

Agentic tools provide a strong context because they can continue work while the user is absent. A prototype could compare return patterns such as a short summary, chronological action history, changed-artifact view, rationale and confidence, or a user-controlled combination. The research would measure not only time to resume but also accuracy of understanding, ability to detect a problematic action, and confidence calibration.

This study would connect the memory-and-resumption literature with mixed initiative without making AI the identity of the thesis. The design problem remains adaptive coordination: what must a system preserve when initiative moves away from and back to the person?

Study 4: Separate sensing quality from interaction quality

The current prototype entangles estimator behaviour with interface response. A future study should hold the adaptation pattern constant while varying signal reliability, or simulate known error rates through Wizard-of-Oz control. This would allow the interface to be evaluated under false positives, false negatives, delayed estimates, and prolonged ambiguous states.

Such a design would reveal whether legibility, reversal, and override actually protect the user when the model fails. It would also prevent a better sensor from being treated as the only path to a better interaction.

Study 5: Revisit physiology only where it adds value

Muse EEG and webcam behaviour were useful for making uncertainty concrete, but future physiological work should follow the interaction question rather than lead it. If physiology is retained, the study should use a frozen build, longer calibration, adaptive baselines, and a larger counterbalanced sample. A research-grade eye tracker could replace the webcam proxy if point-of-gaze is genuinely necessary.

The decision criterion should be whether physiology enables a meaningful adaptive behaviour unavailable through task context or ordinary interaction data. If not, behavioural and contextual signals may be more practical, less intrusive, and easier to explain.

Together, these studies would move the work from "Can an interface react to a workload estimate?" toward "What qualities allow people to coordinate with an adaptive system that may be wrong?" That is the stronger research contribution.

8. Method and Archive Note

This dossier consolidates work developed across two semesters. Its evidence base contains authored research posts, course reflections, a Zotero export of 132 records, 120 locally stored PDFs, detailed annotations prepared for research communication, visual synthesis boards and the complete prototype development record. These materials were not all read or used at the same depth.

The synthesis therefore distinguishes three levels. Close-read sources support theoretical and empirical claims. Bibliography-level records map the wider field and future reading. Project records support claims about what was built, changed and observed. Reflections support claims about the evolution of the researcher's position, not claims about users.

Writing and public explanation were used as research activities. Successive essays exposed conceptual overlap and helped connect older HCI questions with current proactive systems. Presenting the work to a practitioner audience revealed that a broad field map needed a concrete object. The prototype then made assumptions about sensing, thresholds, assistance and override executable.

The development log is part of the evidence. Early runs failed to record useful sensor traces or trigger adaptation reliably. Later revisions added logging, phasic task demand, personal baseline capture and counterbalancing. Those changes improved the probe but divided the pilot into formative and working-build sessions. The dossier reports that limitation rather than pooling the sessions as one stable experiment.

9. Claim Boundaries

The dossier supports the following claims:

- The research developed through a sustained, multi-method inquiry rather than one isolated prototype.
- Interruption cost involves timing, workload, emotion, memory, and recovery.
- A real end-to-end adaptive prototype was built with EEG, webcam-derived behaviour, task demand, and manual control.
- The pilot revealed a plausible reaction-time pattern on the working build alongside lower accuracy, longer resumption lag, and no clear subjective-workload benefit.
- Signal uncertainty and individual variation materially affected when adaptation triggered.
- These observations justify further research into legibility, proportionality, reversibility, and override.

The dossier does not support the following claims:

- That the prototype accurately detected attention or cognitive workload.

- That Muse EEG or webcam behaviour supplied ground truth about mental state.
- That the adaptive condition improved performance overall.
- That findings from four participants generalise to other users or contexts.
- That all 132 library records received equivalent close reading.
- That the current research question is a completed or supervisor-approved thesis definition.

References

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 271-278). ACM. https://doi.org/10.1145/985692.985727

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., Chang, R., & Jacob, R. J. K. (2014). Dynamic difficulty using brain metrics of workload. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3797-3806). ACM. https://doi.org/10.1145/2556288.2557230

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. Cognitive Science, 26(1), 39-83. https://doi.org/10.1207/S15516709COG2601_2

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2015). What makes interruptions disruptive?: A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2971-2980). ACM. https://doi.org/10.1145/2702123.2702156

Bradbury, N. A. (2016). Attention span during lectures: 8 seconds, 10 minutes, or more? Advances in Physiology Education, 40(4), 509-513. https://doi.org/10.1152/advan.00109.2016

Chen, Y., Zhang, C., Fang, W., & Ma, J. (2025). The effects of cues on task interruption recovery in a concurrent multitasking environment. Scientific Reports, 15, 25992. https://doi.org/10.1038/s41598-025-09358-4

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.

Dourish, P. (2004). What we talk about when we talk about context. Personal and Ubiquitous Computing, 8(1), 19-30. https://doi.org/10.1007/s00779-003-0253-8

Epperson, W., Bansal, G., Dibia, V., Fourney, A., Gerrits, J., Zhu, E., & Amershi, S. (2025). Interactive debugging and steering of multi-agent AI systems. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-15). https://doi.org/10.1145/3706598.3713581

Fairclough, S. H. (2009). Fundamentals of physiological computing. Interacting with Computers, 21(1-2), 133-145. https://doi.org/10.1016/j.intcom.2008.10.011

Fitz, N., Kushlev, K., Jagannathan, R., Lewis, T., Paliwal, D., & Ariely, D. (2019). Batching smartphone notifications can improve well-being. Computers in Human Behaviour, 101, 84-94. https://doi.org/10.1016/j.chb.2019.07.016

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. arXiv. https://doi.org/10.48550/arXiv.1311.2222

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 159-166). ACM. https://doi.org/10.1145/302979.303030

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (pp. 1489-1492). ACM. https://doi.org/10.1145/1056808.1056948

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis, and directions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 677-686). ACM. https://doi.org/10.1145/1240624.1240730

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Wozniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. ACM Computing Surveys, 55(13s), Article 286, 1-39. https://doi.org/10.1145/3582272

Koundal, N., Abdalhadi, A., Al-Quraishi, M. S., Elamvazuthi, I., Moosavi, M. S., Guillet, C., Merienne, F., & Saad, N. M. (2024). Effect of interruptions and cognitive demand on mental workload: A critical review. IEEE Access, 12, 54405-54425. https://doi.org/10.1109/ACCESS.2024.3383142

Lavie, N. (2010). Attention, distraction, and cognitive control under load. Current Directions in Psychological Science, 19(3), 143-148. https://doi.org/10.1177/0963721410370295

Li, T., Haines, J. K., de Eguino, M. F. R., Hong, J. I., & Nichols, J. (2022). Alert now or never: Understanding and predicting notification preferences of smartphone users. ACM Transactions on Computer-Human Interaction, 29(5), 1-33. https://doi.org/10.1145/3478868

Liikkanen, L. A., & Gómez, P. G. (2013). Designing interactive systems for the experience of time. In Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces (pp. 146-155). ACM. https://doi.org/10.1145/2513506.2513522

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind?: Examining the nature of fragmented work. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 321-330). ACM. https://doi.org/10.1145/1054972.1055017

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. Human-Computer Interaction, 17(1), 63-139. https://doi.org/10.1207/S15327051HCI1701_2

Mehrotra, A., & Musolesi, M. (2018). Intelligent notification systems: A survey of the state of the art and research challenges. arXiv. https://doi.org/10.48550/arXiv.1711.10171

Monge Roffarello, A., Lukoff, K., & De Russis, L. (2023). Defining and identifying attention capture deceptive designs in digital interfaces. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-19). ACM. https://doi.org/10.1145/3544548.3580729

Monge Roffarello, A., De Russis, L., & Lukoff, K. (2025). The digital attention heuristics: Supporting the user's attention by design. ACM Transactions on Computer-Human Interaction, 32(4), 1-41. https://doi.org/10.1145/3725215

Mozannar, H., Bansal, G., Tan, C., et al. (2025). Magentic-UI: Towards human-in-the-loop agentic systems. arXiv. https://doi.org/10.48550/arXiv.2507.22358

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015). Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In 2015 IEEE International Conference on Pervasive Computing and Communications (pp. 96-104). IEEE. https://doi.org/10.1109/PERCOM.2015.7146515

Pielot, M., & Rello, L. (2017). Productive, anxious, lonely: 24 hours without push notifications. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-11). ACM. https://doi.org/10.1145/3098279.3098526

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology, 40(1-2), 187-195. https://doi.org/10.1016/0301-0511(95)05116-3

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. Cognition and Instruction, 8(4), 351-362. https://doi.org/10.1207/s1532690xci0804_5

Treacy Solovey, E., Afergan, D., Peck, E. M., Hincks, S. W., & Jacob, R. J. K. (2015). Designing implicit interfaces for physiological computing: Guidelines and lessons learned using fNIRS. ACM Transactions on Computer-Human Interaction, 21(6), Article 35, 1-27. https://doi.org/10.1145/2687926

Weber, D., Voit, A., Auda, J., Schneegass, S., & Henze, N. (2018). Snooze!: Investigating the user-defined deferral of mobile notifications. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-13). ACM. https://doi.org/10.1145/3229434.3229436

Zhou, J., Roy, A., Gupta, S., Weitekamp, D., & MacLellan, C. J. (2026). When should users check? Modeling confirmation frequency in multi-step agentic AI tasks. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (pp. 1-20). https://doi.org/10.1145/3772318.3790655

Zindulka, T., Goller, S., Fernandes, D., Welsch, R., & Buschek, D. (2026). The AI memory gap: Users misremember what they created with AI or without. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (pp. 1-22). https://doi.org/10.1145/3772318.3791494