applied sciences

MDPI

Perspective

Cognitive Engagement as a Key Mechanism of Cognitive Sustainability in the Age of Generative AI: An Interdisciplinary Framework

René Riedl [1,2,*] and Frank Krueger [3]

[1] Digital Business Institute, University of Applied Sciences Upper Austria, 4400 Steyr, Austria
[2] Department of Business Informatics—Information Engineering, Johannes Kepler University Linz, 4040 Linz, Austria
[3] School of Systems Biology, George Mason University, Fairfax, VA 22030-4444, USA; fkrueger@gmu.edu
* Correspondence: rene.riedl@jku.at or rene.riedl@fh-steyr.at

Abstract

Generative Artificial Intelligence (GenAI) systems can improve productivity, yet their long-term implications for human cognition remain contested. This article develops an interdisciplinary framework for explaining how different patterns in the distribution and support of cognitive operations between humans and GenAI may shape cognitive engagement and cognitive outcomes over time. The framework draws upon a Neuro-Information-Systems (NeuroIS)-informed approach by integrating cognitive, behavioral, self-report, and neurophysiological perspectives into a multilevel account of how GenAI use shapes human cognition. We argue that cognitive engagement—the depth, persistence, and strategic quality of human processing during AI-supported tasks—is a proposed central mechanism linking the distribution and support of cognitive operations between humans and GenAI to cognitive outcomes. Direct effects on outcomes may coexist with engagement-mediated effects on outcomes. The cognitive consequences depend particularly on whether cognitively consequential operations related to understanding, learning, critical evaluation, and independent judgment continue to involve meaningful human engagement. Specifically, the framework integrates five elements: (1) antecedent conditions (user, task, and GenAI/system characteristics); (2) regulatory processes (metacognition, trust calibration, and motivational/goal regulation); (3) human–AI cognitive work configuration and cognitive engagement (i.e., how cognitive operations are distributed between the user and the AI system and how cognitively engaged the user remains); (4) short- and intermediate-term outcomes (performance, understanding, retention, and agency); and (5) long-term outcomes of cognitive sustainability and cognitive dependence. The article derives design principles for engagement-preserving AI and contributes to an interdisciplinary research agenda for examining these relationships through convergent behavioral, self-report, neurophysiological, and longitudinal evidence.

Keywords: generative AI (GenAI); cognitive sustainability; cognitive offloading; cognitive scaffolding; cognitive engagement; neuro-information-systems (NeuroIS); neuroergonomics; human–AI interaction; metacognition; trust calibration; sustainable learning

Check for updates

Academic Editor: Jing Jin

Received: 5 October 2026
Revised: 28 October 2026
Accepted: 29 October 2026
Published: 3 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Humans have always expanded their minds beyond the brain. From calculators to search engines, we have continuously developed tools that reduce cognitive effort. This phenomenon, known as cognitive offloading, is not new; it is a cornerstone of human progress.

Research on cognitive offloading, external memory, cognitive artifacts, transactive memory, and distributed cognition has long shown that human thinking is routinely organized through tools, environments, and social systems rather than being confined to the individual brain [1–7]. However, something fundamental has changed with Generative Artificial Intelligence (GenAI): we are no longer only outsourcing storage, navigation, calculation, or information retrieval. We are increasingly outsourcing drafting, summarizing, explaining, coding, recommending, and even, at least in part, judgment and decision-making.

This shift gives rise to a new version of an old question: Do digital tools free up cognitive resources for higher-level thinking, or do they replace the processes through which people learn, remember, reason and develop expertise? The question has become especially urgent because GenAI can produce fluent outputs that look like understanding, even when the human user has not engaged deeply with the underlying content. Empirical studies already show that GenAI can improve productivity and task performance in some settings [8–10], while other work raises concerns about reduced critical thinking, overreliance, weaker ownership, and shallower processing [11,12]. The difficulty is that both claims can be true. GenAI can amplify cognition in one interaction and weaken cognitive engagement in another.

Against this background, the focus of this article is not on whether GenAI is good or bad for human cognition. This question is too simplistic. Instead, it asks how cognitive operations are distributed between humans and GenAI systems and under what conditions patterns of offloading and scaffolding preserve or reduce meaningful human cognitive engagement. We propose cognitive engagement as a central, but not exclusive, mechanism. Offloading and scaffolding may nevertheless co-occur when different operations are delegated, retained, or supported within the same task. When AI interaction suppresses retrieval, explanation, verification, and judgment, it shifts users toward offloading. When AI interaction prompts attempts, reflection, critique, revision, and metacognitive monitoring, it can function as scaffolding.

The contribution of this article is to develop a NeuroIS-informed conceptual framework explaining how different patterns in the distribution of cognitive operations between humans and GenAI shape cognitive engagement and may contribute to cognitive sustainability or cognitive dependence over time. Neuro-Information-Systems (NeuroIS) is a research discipline that relies on neuroscience and neurophysiological knowledge and tools to better understand the development, use, and impact of information and communication technologies [13,14]. A defining feature of NeuroIS is the combined use of behavioral measures (e.g., clicking, navigation, task performance), self-reports (e.g., perceived usefulness, trust, workload, satisfaction), and neurophysiological measures (e.g., electroencephalography, EEG; functional magnetic resonance imaging, fMRI; functional near-infrared spectroscopy, fNIRS, eye tracking, heart rate, or electrodermal activity) [15,16]. These complementary data sources capture what users do, what they consciously report, and how their brain and body respond during technology use, while longitudinal measurements may help examine whether these patterns are associated with persistent changes in neural structure and cognitive functioning over time. Their integration provides a richer understanding of user cognition, emotion, and behavior, while also informing information systems design [16,17]. The proposed framework links GenAI system, user, and task characteristics; trust calibration, metacognitive regulation, and motivational/goal regulation; the distribution of cognitive operations and cognitive engagement; short- and intermediate-term cognitive outcomes; and the potential long-term outcomes of cognitive sustainability and cognitive dependence.

In this article, we define *cognitive sustainability* as the long-term preservation, renewal, and development of the cognitive capacities and agency required for individuals to think, learn, judge, and act effectively in AI-mediated environments. The concept is used as an

integrative framing lens rather than as a fully mature standardized psychological construct. It shifts the focus from short-term productivity and task performance to whether repeated AI-supported interactions preserve and develop the human capacities required for effective learning, judgment, and action over time [18,19].

Moreover, we use *cognitive dependence* to describe a potential long-term outcome in which repeated reliance on AI increasingly substitutes for cognitive operations that users would otherwise need to perform themselves, resulting in reduced capacity to perform relevant tasks independently and/or reduced willingness to do so when equivalent AI assistance is unavailable. Cognitive dependence should not be understood as a clinical construct, as simply the opposite pole of cognitive sustainability, or as an inevitable consequence of cognitive offloading; rather, it represents a theoretically proposed long-term risk when repeated patterns of AI reliance displace opportunities to maintain or develop relevant cognitive capabilities. Accordingly, cognitive sustainability and cognitive dependence represent theoretically distinct potential long-term outcomes rather than opposite ends of a single continuum. Because reduced capacity and reduced willingness are conceptually distinct, future studies should assess them separately.

This article makes four contributions. First, it shifts the focus of the debate about GenAI and cognition from the question of whether AI is helpful or harmful to the question of how cognitive operations are allocated: which ones remain with the human and which ones are delegated to the system? Second, it proposes cognitive engagement as a partial mediator that links the distribution and support of cognitive operations between humans and GenAI to cognitive outcomes. Third, it integrates trust calibration, metacognitive regulation, and motivational/goal regulation as proximal regulatory conditions shaping whether users verify, challenge, revise, or defer to AI output. Finally, it outlines an interdisciplinary research agenda for measuring these pathways through convergent behavioral, self-report, and neurophysiological indicators.

The remainder of this article is structured as follows. Section 2 introduces the theoretical foundations of human–GenAI cognitive work, focusing on cognitive offloading, cognitive engagement, and cognitive scaffolding. Section 3 develops the interdisciplinary framework of cognitive sustainability in GenAI use and outlines the role of NeuroIS in investigating its core mechanisms and outcomes. Section 4 derives design principles for engagement-preserving AI. Section 5 presents directions for future research, formulates the framework's central propositions, discusses the current evidence base and boundary conditions, and provides an illustrative empirical test. Section 6 discusses the limitations and concludes the article.

2. Theoretical Foundations of Human–GenAI Cognitive Work

2.1. Cognitive Offloading

Cognitive offloading allows individuals to reduce mental effort by relying on external tools. Classic examples include taking notes, using GPS navigation, saving files, or searching the web. Decades of research have shown that this strategy can improve efficiency because it reduces the demands placed on internal memory, attention, and executive control [1,20–22]. At the same time, when users expect information to remain externally available, they may encode less deeply, rely more on recognition than recall, and remember where to find information rather than the information itself [20].

The "Google effects on memory" study illustrates this point. Across four studies, Sparrow, Liu, and Wegner showed that internet access can function as a form of external or transactive memory system: when people expect future access to information, they remember less of the content itself and more about where it can be found [20]. This finding

does not imply that search engines made people less intelligent. Rather, it shows that cognition reorganizes around available tools.

Related findings in navigation show that habitual GPS reliance can be associated with weaker spatial-memory engagement, while memory research shows that saving information can free resources for subsequent learning, but also changes what is retained internally [21,22]. Offloading is therefore not inherently problematic. It allows humans to allocate cognitive resources more strategically. The key question is not whether cognition is offloaded, but which cognitive processes are offloaded.

Offloading low-value memory burden may be adaptive. Offloading the very operations through which learning occurs—retrieval, elaboration, self-explanation, verification, comparison, and judgment—may be cognitively costly. This distinction is especially important for GenAI because such systems do not merely store information for later use. They can produce the answer, the structure, the argument, the summary, the code, the critique, and the next step. As a result, cognitive offloading expands from external memory into externalized symbolic production and judgment-like activity.

2.2. The AI Shift: From External Tool to Cognitive Agent

GenAI fundamentally transforms this dynamic. Unlike earlier technologies, such systems do not simply store or retrieve knowledge. They generate new content, which still requires evaluation. A notebook stores traces; a calculator transforms numbers; a search engine retrieves candidate sources. GenAI, in contrast, can synthesize, translate, simplify, elaborate, design, simulate, and persuade. It can therefore contribute to the task not merely as a resource, but as a quasi-collaborator actively involved in generating cognitive output.

Recent work from the MIT Media Lab, widely discussed under the title "Your Brain on ChatGPT," offers early insights into how this shift might affect cognitive engagement. In an essay-writing task, participants using a large language model (LLM) showed, based on EEG, lower EEG-derived connectivity during the task, lower reported ownership of their essays, and poorer recall of their own text than participants writing without tools or with search-engine support [12]. These findings are suggestive and important, but they should be interpreted cautiously: the study is a preprint, the sample is modest, and the task is specific. The strongest interpretation is not that GenAI necessarily weakens cognition, but that answer-first AI support may reduce engagement under some conditions. This is precisely why a NeuroIS research agenda is needed.

The broader empirical picture is heterogeneous. GenAI has been shown to improve productivity in professional writing and customer-support work [8,9]. Field experiments with consultants suggest a "jagged technological frontier": AI can improve performance on tasks within the model's capability frontier, while harming performance on tasks outside it [10]. A survey of knowledge workers found that higher confidence in AI was associated with lower self-reported critical-thinking effort, whereas higher task-specific self-confidence was associated with greater reported critical thinking [11]. These associations do not establish causation, and self-confidence is not equivalent to objectively measured expertise. In high-school mathematics, unguarded AI support improved assisted practice performance, but reduced subsequent unaided exam performance; tutoring safeguards largely mitigated this penalty rather than establishing superior unaided learning [23]. Work on AI-generated code likewise highlights design-dependent trade-offs among engagement, friction, and calibration [24].

The implication is that GenAI should not be studied as a single treatment, but rather as an environment that can be designed for interaction. The same system can facilitate learning in one context and hinder it in another. This motivates cognitive engagement as a

candidate explanatory construct for why immediate performance gains may coexist with weaker retention, ownership, calibration, or future independent performance.

2.3. Cognitive Engagement as a Proposed Central Mechanism

Building on educational-psychology accounts of cognitive engagement and active learning, we define *cognitive engagement* in GenAI use as the depth, persistence, and strategic quality of human cognitive processing during AI-assisted task performance [25–27]. This definition is the authors' synthesis adapted to AI-mediated work and learning. It includes interpreting, retrieving, elaborating, comparing, explaining, verifying, revising, and taking responsibility for final judgment.

This definition is deliberately narrower than general "engagement" and differs from related constructs. Cognitive effort and cognitive load concern the amount of mental resources expended or the demands imposed by the task [28]. Attention concerns the allocation of perceptual and mental resources [29]. User engagement in human–computer interaction includes broader experiential dimensions such as focused attention, novelty, involvement, aesthetics, and challenge [30]. Flow refers to immersive, intrinsically rewarding absorption [31]. Metacognition and self-regulated learning concern monitoring and control processes that help organize learning activity [32–34]. Cognitive engagement, by contrast, concerns the quality of task-directed processing itself: whether the user is making sense, checking, elaborating, revising, and retaining responsibility for judgment. A user may pay attention to AI output without evaluating it; may expend effort on a shallow task; may enjoy an interaction without learning; or may rely on metacognitive monitoring without actually processing the content deeply. High cognitive engagement requires meaningful processing of content and responsibility for sense-making.

Specific learning strategies, such as retrieval practice and self-explanation, are possible ways of enacting cognitive engagement; their presence does not establish processing quality. Offloading and scaffolding describe how cognitive operations are delegated or supported, whereas engagement concerns the human processing within those configurations. Empirical tests should therefore distinguish the support provided, independently assessed engagement, and subsequent outcomes. A required explanation can be superficial or copied. Equally, a short interaction can reflect efficient, substantive processing rather than disengagement.

The construct has strong roots in educational psychology. Corno and Mandinach described cognitive engagement as strategic, effortful processing for learning [25]. Fredricks and colleagues situated cognitive engagement within a broader engagement taxonomy that distinguishes cognitive from behavioral and emotional engagement [26]. Chi and Wylie's ICAP framework further shows that learning outcomes depend on whether learners are interactive, constructive, active, or passive [27]. In ICAP, passive engagement involves receiving information, active engagement involves manipulating provided material, constructive engagement involves generating new inferences or explanations, and interactive engagement involves co-constructive dialogue with another agent. These distinctions are directly relevant to GenAI. Reading an AI answer is often passive. Copying or lightly editing it may be active. Explaining, critiquing, comparing, or revising it may become constructive. Dialoguing with the system to test ideas may become interactive.

Human–AI research provides an especially relevant bridge. In nutritional decision tasks using simulated AI assistance, Gajos and Mamykina found that recommendations with explanations improved immediate decisions but did not produce learning gains; explanation-only support that required participants to derive their own decision improved both decisions and learning [35]. These task-specific findings support distinguishing as-

sisted performance from learning. They do not establish long-term cognitive sustainability or validate engagement as a mediator, which remains a hypothesis in the present framework.

Cognitive engagement therefore helps explain the cognitive consequences of different configurations of cognitive offloading and cognitive scaffolding. These patterns are not mutually exclusive and may co-occur within the same task. Offloading may reduce cognitive engagement when AI reduces the need for human processing of cognitively consequential task operations. Scaffolding may preserve or increase cognitive engagement when AI reduces unproductive burden while supporting the operations that matter for learning, judgment, and transfer.

2.4. Cognitive Scaffolding

Cognitive scaffolding is well established in educational psychology, but underexplored in several related domains, including cognitive neuroscience, neuropsychology, and NeuroIS. Scaffolding provides support that enables individuals to perform tasks they could not complete independently, while still requiring active mental participation. The classical roots of this idea lie in Vygotsky's zone of proximal development and in Wood, Bruner, and Ross's analysis of tutoring [36,37]. Later work clarified that scaffolding is not simply help. It requires contingent support, fading, and transfer of responsibility back to the learner [38].

This distinction is particularly important for understanding GenAI use and its possible consequences. A system that gives a complete answer is immediately helpful, but it is not necessarily a scaffold. A scaffold structures the learner's activity. It may break a task into manageable steps, ask the user to predict before revealing, provide hints instead of solutions, require self-explanation, invite comparison of alternatives, or prompt reflection after revision. Software-scaffolding research emphasizes that effective scaffolds both structure the task and problematize the important conceptual work, so that support reduces unproductive burden without eliminating sense-making [39,40]. In this sense, cognitive scaffolding is a design principle, not merely a benevolent intention.

Learning science provides several mechanisms that are particularly relevant to scaffolded GenAI. Self-explanation helps learners integrate new information with prior knowledge and improves understanding when it is actively elicited [41]. Retrieval practice strengthens later recall and transfer by requiring learners to reconstruct knowledge rather than merely recognize it [42,43]. Productive failure can support durable learning when initial struggle is followed by appropriate consolidation and feedback [44]. Metacognitive monitoring helps users decide whether they understand or merely recognize fluent output [32–34].

In AI-assisted learning and work, scaffolding therefore means that the user remains in the loop of meaning-making. The AI may help formulate questions, identify gaps, suggest counterarguments, or provide feedback, but the human must still generate, interpret, verify, and revise. The desired endpoint is not permanent dependence on the AI. Rather, it is increasing human competence, calibration, and agency.

3. An Interdisciplinary Framework of Cognitive Sustainability in GenAI Use

The framework responds to the problem identified in Section 1: improvements in immediate AI-supported performance need not imply equivalent benefits for human learning or independent capability [8–12]. The synthesis connects that problem to complementary accounts of delegation, activity-preserving support, processing quality, and regulation.

3.1. Conceptual Synthesis Approach

This Perspective article uses a theory-guided narrative conceptual synthesis, not a systematic review. The organizing question is how delegating and supporting cognitive operations can produce different consequences for human processing and capability. The traditions discussed in Sections 2.1–2.4 and 3.2 were selected for complementary explanatory roles: offloading describes delegation; scaffolding describes contingent support; engagement specifies processing quality; metacognition and trust address regulation; and NeuroIS informs multimethod investigation. Foundational accounts and task-specific empirical studies are used to define, motivate, or qualify these roles, rather than to represent an exhaustive evidence inventory. The integration connects conditions of use to regulation, operation-level work configuration, engagement, and outcomes at different timescales. With reference to Section 5.1, in which we describe major propositions that result from the interdisciplinary framework, this reasoning motivates P1–P3 on configuration and cognitive engagement, P4–P5 on regulatory processes, P6 on contextual variation, P7 on multi-method measurement, and P8 on repeated-use consequences. The framework was refined through conceptual comparison and manuscript revision; it was not inductively extracted from a predefined, systematically screened corpus. The proposed ordering and relationships are the authors' theoretical synthesis, not findings established collectively by the cited studies. In Section 5.2, we distinguish these inferences from empirical evidence, and the non-exhaustive scope remains a limitation.

3.2. Cognitive Offloading and Cognitive Scaffolding

Against this background, the perspective presented in this article distinguishes different patterns in how people distribute cognitive work between themselves and AI systems, with potentially distinct cognitive consequences. This distinction should not be read as a simple good-versus-bad dichotomy. Offloading can be useful, and scaffolding can be poorly designed. The analytical distinction in Section 2 therefore concerns the operations delegated or supported, not a classification of users or tasks as inherently engaged or disengaged.

In cognitive offloading, particular cognitive operations are delegated to AI, which may act as a task executor for those operations. Users may take on a more passive role when they accept or copy AI-generated outputs without much mental processing. Although this approach reduces effort and may increase efficiency, it is associated with a risk of more superficial processing, reduced learning, weaker long-term retention, and lower task ownership when AI performs the processes users need to practice [11,12,35]. The proposed risk is particularly relevant when AI performs retrieval, explanation, checking, weighing alternatives, or forming independent judgments on behalf of the user.

Cognitive scaffolding, which may coexist with offloading within the same interaction, supports a more active and constructive use of AI. In this approach, GenAI acts as a supporter or coach, helping users structure their thinking, refine their ideas, and engage in iterative reasoning. Users remain cognitively involved by providing initial attempts, asking follow-up questions, evaluating AI-generated responses, and integrating new information into existing knowledge structures. The GenAI tool reduces unnecessary burden but preserves the cognitive operations through which understanding develops.

To clarify this analytical distinction, Table 1 contrasts offloading and scaffolding across their core function, AI role, user role, engagement profile, learning consequences, risks, and appropriate use cases [1,27,35,37,38,45].

Table 1. Cognitive offloading and cognitive scaffolding in GenAI use.

Dimension	Cognitive Offloading	Cognitive Scaffolding
Core function	Transfers specific cognitive operations from human to AI	Structures and supports cognitive operations while preserving active human involvement
Typical AI role	Task executor, answer producer, drafter, recommender	Coach, tutor, critic, hint provider, reflective partner
Typical user role	Delegator of selected cognitive operations; may remain actively involved in others	Generator, explainer, verifier, reviser, judgment steward
Cognitive engagement	Varies with the type and extent of operations delegated; may decline when cognitively consequential operations are offloaded	Intended to be sustained through attempts, explanation, critique, and revision; actual processing quality must be assessed independently
Learning consequence	May improve immediate output while weakening retention or transfer when cognitively consequential operations are repeatedly delegated	May support retention, transfer, calibration, and agency when support elicits meaningful processing
Main risk	Dependence, false fluency, shallow understanding, reduced ownership	Over-scaffolding, frustration, excessive friction, or failure to fade support
Best use case	Routine, low-stakes tasks where learning is not the goal and verification is easy	Learning, high-stakes judgment, professional development, and complex problem solving

Note: Cognitive offloading and cognitive scaffolding are analytical patterns rather than mutually exclusive task types; they may co-occur within the same AI-supported interaction across different cognitive operations, task stages, and degrees of delegation.

Timing is one design variable rather than a sufficient condition for learning. An initial human attempt may help preserve involvement in subsequent AI-supported reasoning, but Gajos and Mamykina also observed improved decisions without learning when participants chose before seeing an AI recommendation and explanation [35]. Attempt-first interaction should therefore be evaluated together with the processing required after assistance, its fit to prior knowledge, and the quality of the user's reasoning, rather than treated as a guarantee of engagement or retention.

Granularity also matters: AI can provide full answers, partial suggestions, guiding questions, explanations, feedback, or prompts for reflection. The more complete and ready-to-use the output, the greater the risk of offloading cognitively consequential operations. The more the system requires the user to think, compare, and decide, the greater the potential for cognitive scaffolding that preserves active human involvement. This is why "better AI" is not always better for cognition. A more fluent answer can reduce the need for human processing unless the interface deliberately preserves engagement.

User intent is another decisive factor. The same AI-generated output can be used in different ways. A student may copy an answer or use it to identify gaps in understanding. A professional may accept a recommendation or use it as a stimulus for further analysis. However, user intent is not sufficient. Interface design, incentives, time pressure, assessment design, accountability, and organizational norms all shape how cognitive operations are distributed between the human and the AI system and whether users remain meaningfully cognitively engaged.

Three regulatory processes are especially important: trust calibration, metacognitive regulation, and motivational/goal regulation. Trust calibration means that reliance on AI should track actual system trustworthiness in a particular task context [46–49]. Overtrust

promotes passive acceptance and automation bias, whereas undertrust leads to disuse and missed opportunities for complementarity [47,50,51]. Metacognitive regulation means that users monitor their understanding, confidence, uncertainty, and need for verification [32–34,52]. Motivational/goal regulation concerns the goals users pursue during AI-supported task performance and the extent to which they regulate their effort toward learning, understanding, or immediate task completion. Learning-oriented goals are more likely to preserve engagement in cognitively consequential operations, whereas a predominant focus on rapid task completion may increase their delegation to AI. Together, these regulatory processes shape how cognitive operations are distributed between the human and the AI system and whether users remain meaningfully engaged in evaluating, verifying, and integrating AI-generated information.

Social and organizational psychology further specifies the context of these processes. Transactive-memory research examines how groups distribute expertise and coordinate access to what members know [4,53]. Social facilitation and inhibition show that the presence of others can affect performance differently depending on task demands [54]; accountability research likewise distinguishes conditions under which evaluation improves judgment from those under which it preserves or amplifies bias [55]. These findings concern human social settings, not direct tests of GenAI use. Applied here as theoretical extensions, they suggest that coordination demands, evaluation arrangements, shared responsibility, and incentives may shape whose contributions are checked, when uncertainty is disclosed, and whether users prioritize understanding or rapid completion. In the existing framework, these are contextual antecedents that may influence trust calibration and metacognitive regulation (P4–P5), qualify the effects of AI support (P6), and motivate measuring social context alongside behavioral and physiological profiles (P7). The extension does not assume that an AI system is psychologically equivalent to a human collaborator or that social influence invariably benefits engagement.

The balance between offloading and scaffolding can thus be analyzed at the level of individual cognitive operations. For each operation, researchers should distinguish who performs it, how support is delivered, and the quality of the human contribution, rather than infer engagement from the mere availability of AI assistance.

Timing, granularity, user intent, and the regulatory processes described above identify candidate mechanisms and boundary conditions. Their effects should be evaluated within the task and social context, rather than assuming that a particular interface necessarily preserves engagement.

This operation-level account guides the model in Section 3.3 and the empirical illustration in Section 5.3. It accommodates selective delegation while retaining independent measures of processing quality and later capability.

3.3. NeuroIS Approach to Investigating Cognitive Sustainability

NeuroIS provides a complementary methodological approach to studying this framework by combining behavioral measures, self-reports, and neurophysiological measurements [14,16]. Behavioral data, such as interaction logs, navigation behavior, verification actions, revision traces, and task performance, reveal what users do during AI interaction. Self-reports capture conscious perceptions such as trust, perceived understanding, cognitive effort, confidence, mental workload, and task ownership. Neurophysiological measures, such as EEG, fMRI, fNIRS, eye tracking, pupillometry, electrodermal activity, and heart-rate variability, provide indicators of psychological and physiological processes that may not be accessible through behavioral observation or introspection alone [15]. Longitudinal designs may extend these measures through repeated behavioral, self-report, EEG, and neuroimaging assessments to examine whether patterns of AI-supported cognitive work

are associated with persistent changes in cognitive functioning. Because relevant time scales are likely to vary by task, population, and outcome, the framework does not impose fixed thresholds; empirical studies should instead prespecify exposure and follow-up windows. By integrating complementary data sources across timescales, NeuroIS can examine how patterns of cognitive offloading and scaffolding relate to cognitive engagement and to short-, intermediate-, and potential long-term outcomes.

3.4. Interdisciplinary Framework of Cognitive Sustainability in GenAI Use

Conceptually, the proposed framework connects antecedent conditions to potential long-term cognitive outcomes through regulatory processes, the distribution of cognitive operations between the human and the AI system, cognitive engagement, and short/intermediate-term outcomes. Figure 1 summarizes this framework as five analytically distinct but interrelated components. The first component comprises antecedent conditions related to the user, task, and GenAI system. User factors include expertise and prior knowledge, AI literacy, motivation and goals, and cognitive capacity. Task factors include complexity and novelty, time pressure, accountability and stakes, and learning versus performance goals. Social and organizational conditions—including coordination demands, evaluation arrangements, and incentives—contextualize these factors. GenAI/system factors include answer-first versus scaffolded interaction design, autonomy and adaptivity, transparency and uncertainty communication, source visibility, and friction and fading. These antecedent conditions are proposed to shape regulatory processes and cognitive work distribution; direct influences on outcomes and additional pathways are not excluded.

Three proximal regulatory processes play a particularly important role. Metacognitive regulation determines whether users monitor their understanding, recognize uncertainty, evaluate their confidence, and decide when additional verification or revision is required. Trust calibration shapes whether and to what extent users rely on AI, verify outputs selectively, or accept recommendations uncritically. Motivational and goal regulation influences the willingness to invest cognitive effort and whether users prioritize learning, understanding, and skill development or immediate task completion and performance. Together, these regulatory processes shape how cognitive operations are distributed between the human and the AI system and the quality of cognitive engagement within these patterns.

With reference to Section 5.1, cognitive engagement is proposed as a partial mediator between human–AI cognitive work configuration and short- and intermediate-term cognitive outcomes (P1). This permits a direct route from configuration to the short- and intermediate-term outcomes such as task performance alongside the indirect route through cognitive engagement. For example, delegating drafting may improve the report produced without increasing the user's understanding, whereas evaluating and integrating evidence may contribute to later unaided performance. The framework does not require a nonzero direct effect for every outcome or exclude other mediators. Which paths operate, and their strength, are empirical questions rather than established causal results.

Selective delegation is compatible with this partial-mediation account. Offloading routine operations may reduce unnecessary burden while leaving verification, comparison, integration, or judgment with the user. Conversely, repeated delegation of the operations being learned may reduce practice opportunities even when the assisted product is accurate. The distinction in Section 2 therefore becomes a prediction about operation-specific effects, not a universal ranking of more versus less AI use.

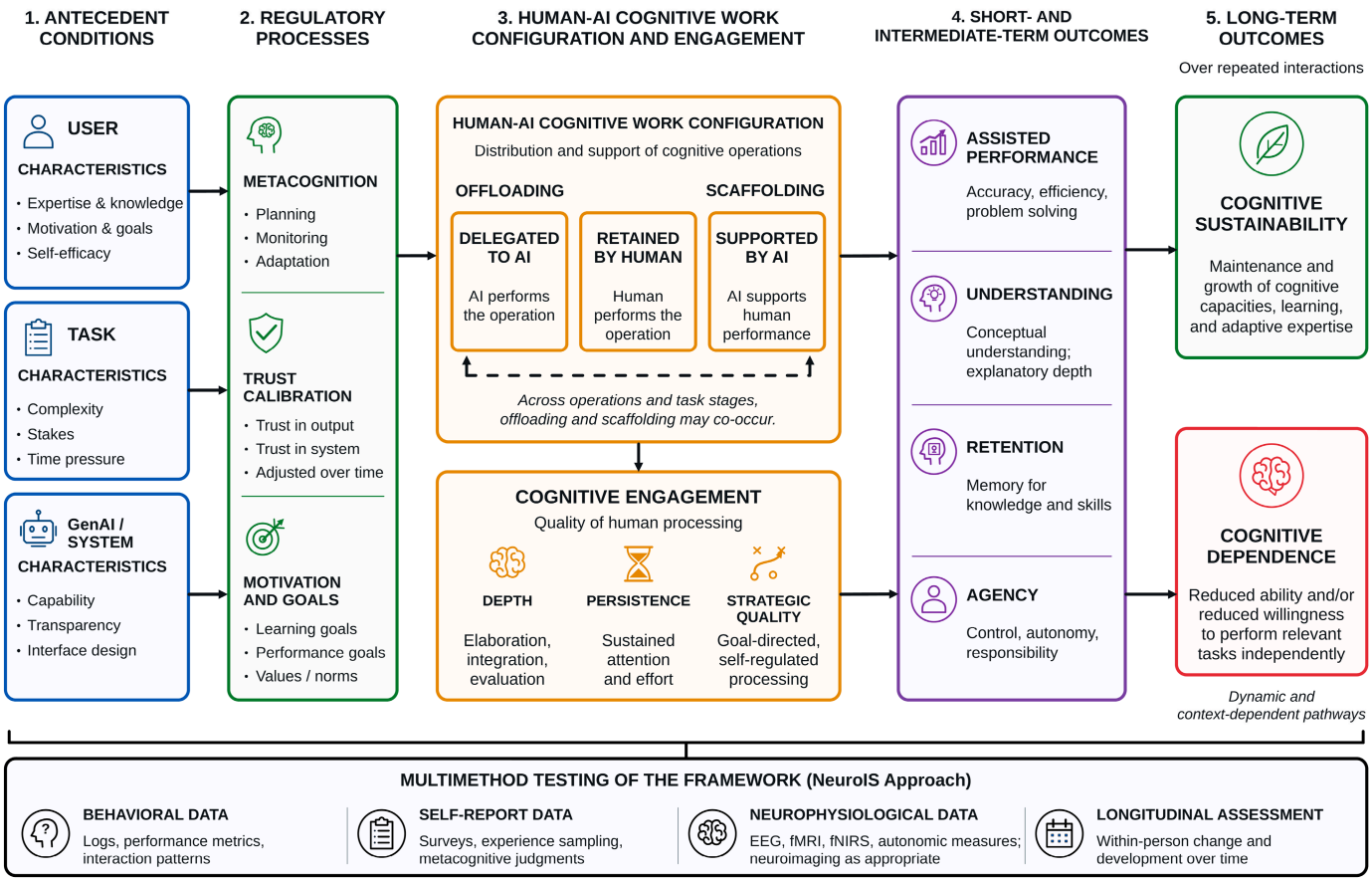


Figure 1. Interdisciplinary framework of cognitive sustainability in GenAI use. Note: The proposed framework comprises five interconnected components, presented from left to right. (1) Antecedent conditions include user characteristics (expertise and knowledge, motivation and goals, and self-efficacy), task characteristics (complexity, stakes, and time pressure), and GenAI/system characteristics (capability, transparency, and interface design). These characteristics are proposed to influence (2) regulatory processes, comprising metacognition (planning, monitoring, and adaptation), trust calibration (trust in AI output and the system, adjusted over time), and motivation and goals (learning goals, performance goals, and values/norms). These processes are proposed to shape (3) human–AI cognitive work configuration and cognitive engagement. Cognitive work configuration distinguishes operations delegated to AI, retained by humans, or supported by AI, acknowledging that cognitive offloading and scaffolding may coexist across operations and task stages. Cognitive engagement refers to the quality of human processing, characterized by depth (elaboration, integration, and evaluation), persistence (sustained attention and effort), and strategic quality (goal-directed, self-regulated processing). The framework proposes that cognitive work configuration may influence (4) short- and intermediate-term outcomes, including assisted performance (accuracy, efficiency, and problem solving), understanding (conceptual understanding and explanatory depth), retention (memory for knowledge and skills), and agency (control, autonomy, and responsibility). Cognitive engagement is proposed as a partial mediator of these relationships, while direct effects of cognitive work configuration on outcomes may also occur. Across repeated interactions, these outcomes may contribute to (5) cognitive sustainability (maintenance and growth of cognitive capacities, learning, and adaptive expertise) or cognitive dependence (reduced ability and/or reduced willingness to perform relevant tasks independently). These are conceptualized as distinct potential long-term outcomes rather than opposite ends of a single continuum. The bottom panel illustrates a complementary multimethod approach to investigating the framework (NeuroIS approach), integrating behavioral data, self-report data, neurophysiological data, and longitudinal assessment. All depicted relationships represent theoretical propositions rather than established causal effects; their direction, magnitude, and boundary conditions require empirical investigation.

Scaffolding is specified by observable support features, including contingent hints, feedback, and fading, rather than by successful learning after the event. These features are intended to support human processing but can fail to do so. Cognitive engagement and outcomes must therefore be assessed independently of whether the interface was designated as scaffolded. Offloading and scaffolding can coexist across operations and stages within the same task.

Several neurocognitive processes are particularly relevant for understanding these pathways. Attention allocation and cognitive control influence whether users inspect AI-generated information selectively, inhibit premature acceptance, compare alternatives, and revise responses [29,56]. Memory encoding and retrieval determine whether AI-supported activity produces durable learning rather than temporary task completion [42,43]. Error monitoring becomes critical whenever AI suggestions conflict with existing knowledge or available evidence [57], while metacognitive monitoring helps users distinguish genuine understanding from the mere recognition of fluent AI-generated responses [32–34,52]. NeuroIS methods cannot directly observe thinking itself; however, they can triangulate these processes by combining neurophysiological measurements with behavioral evidence, self-reports, task performance, and delayed learning outcomes [14,16].

The core behavioral predictions do not require neurophysiological measurement. NeuroIS can add a temporally resolved, convergent account of processes such as error monitoring when AI suggestions conflict with task evidence [14–16,57]. A relevant extension could compare responses to matched, time-locked suggestions while measuring verification and later unaided performance. Lower activation or connectivity alone would not distinguish efficient processing from disengagement. Any brain-related or autonomic indicator requires task-specific validation and control of alternative explanations such as difficulty or arousal; its added value should be demonstrated beyond behavioral and self-report measures.

The patterns of cognitive work distribution shown in Figure 1 also generate different empirical predictions. Interactions characterized by greater offloading of cognitively consequential operations may be associated with less substantive verification and revision and weaker delayed recall or transfer. In contrast, interactions that scaffold and preserve human involvement in cognitively consequential operations are predicted to support better-quality reasoning, verification, and revision, with possible benefits for delayed retention or transfer. Counts of clicks, explanations, or time spent are process descriptors, not sufficient measures of engagement quality. These predictions depend on which cognitive operations are delegated, retained, or supported; selective offloading of routine operations may reduce unnecessary cognitive burden without reducing meaningful cognitive engagement. Neurophysiological indicators should therefore not be interpreted as direct readouts of cognition. Rather, they should be triangulated with behavioral measures, self-reports, and task outcomes to distinguish reductions in unproductive cognitive burden from reductions in meaningful cognitive engagement [16].

Cognitive sustainability and cognitive dependence are modeled as distinct potential long-term outcomes of repeated interaction, without assuming that either directly causes the other. They need not be mutually exclusive across different skills or tasks. Testing dependence requires separating a change in unaided capability from a change in willingness to work independently; preference for a convenient tool alone does not demonstrate cognitive loss. Repeated assessments are needed to distinguish persistent changes from temporary assistance effects.

4. Design Principles for Engagement-Preserving AI

Design has been part of the NeuroIS agenda from the field's inception [13,58]. Rather than viewing neuroscience solely as a means to understand technology use and its conse-

quences, NeuroIS has also emphasized its potential to inform the design and evaluation of information systems. This design perspective is particularly important in the context of AI, where the challenge is not merely to maximize immediate task performance, but to design systems that foster cognitive engagement and long-term cognitive sustainability while limiting the risk of cognitive dependence.

Achieving such design requires a shift in how AI systems are evaluated. Rather than focusing solely on output quality and speed, AI systems should also be assessed by whether they encourage meaningful cognitive engagement. This may involve deliberately postponing full automation, prompting users to think before receiving answers, and encouraging reflection instead of passive acceptance. Human-centered AI guidelines and interaction principles already point toward reliable, understandable, and controllable systems [59,60]. A cognitive-sustainability perspective extends this logic by asking whether the system also preserves and develops human cognitive capabilities over time.

These proposed mechanisms motivate design principles grounded in cognitive engagement, scaffolding, self-explanation, cognitive forcing, and human–AI interaction research [27,37,40,41,45,50,60,61], as summarized in Table 2.

Table 2. Design principles for cognitively sustainable GenAI interaction.

Design Principle	Implementation	Intended Cognitive Function
Attempt-first interaction	Ask users to generate an initial answer, plan, diagnosis, argument, or prediction before receiving full AI support.	May reduce immediate delegation and anchor AI use in prior human cognition; an initial attempt alone does not ensure learning.
Hint ladders	Provide graduated hints before revealing complete answers.	Reduces unnecessary burden while preserving problem-solving activity.
Self-explanation prompts	Require users to explain why an answer is plausible or how it connects to prior knowledge.	Supports elaboration, understanding, and memory encoding.
Retrieval and revision loops	Ask users to recall, compare, revise, and explain what changed after AI feedback.	Converts AI use into practice rather than mere consumption.
Uncertainty and source visibility	Show uncertainty, evidence, alternative views, and verification paths.	Supports trust calibration and reduces passive overreliance.
Adaptive fading	Gradually reduce support as competence increases.	Increases human responsibility as competence develops.
Selective friction	Insert pauses, prompts, or justification requirements in high-stakes or high-uncertainty moments.	Slows reflexive acceptance while avoiding blanket inefficiency.

These principles do not imply that every AI interaction should be difficult. Routine administrative tasks may appropriately be automated. The point is selective: when the goal is learning, expertise development, judgment quality, or long-term human capability, design should preserve the cognitive operations that build those capacities. Their effectiveness must be assessed through independent measures of processing and learning, not inferred from compliance with the interface prompts.

5. Research Agenda

5.1. Directions for Future Research and Propositions

NeuroIS is particularly well positioned to explore the cognitive and neurophysiological consequences of AI use. NeuroIS researchers have investigated human interaction with

information and communication technologies by integrating neurophysiological measures with introspective and behavioral data [14,16,17]. This multimethod approach enables researchers to examine how different patterns of cognitive work distribution and engagement during AI interaction are associated with brain activity, autonomic responses, reliance behavior, cognitive engagement, and learning outcomes. The most defensible near-term agenda is not to claim proven neural harm resulting from GenAI use, but rather to develop theoretically informed, multimethod studies that examine different patterns of AI-supported cognitive work and how they shape cognitive engagement and cognitive outcomes over time.

Individual differences must also be considered. Users vary in expertise, prior knowledge, motivation, AI literacy, cognitive style, personality, age, and metacognitive ability. Personality-related differences have already been discussed in the context of trust in AI systems [48,62,63]. Furthermore, individuals with a high need for cognition (i.e., a strong motivation to engage in demanding mental activities) derived the greatest benefit from cognitive forcing interventions in human–AI interactions [50]. Expertise is likely to be particularly important. Novices may benefit substantially from structured scaffolding, but may also be more vulnerable to accepting fluent AI output without sufficient verification. Experts, by contrast, may be better able to detect errors, integrate AI-generated suggestions into existing knowledge structures, and selectively rely on AI only when appropriate.

The framework's incremental contribution is to connect operation-level delegation and support to processing quality, regulation, and repeated-use outcomes. ICAP characterizes modes of engagement [27], self-regulated learning addresses monitoring and control [34], and human–AI research examines reliance and interaction design [50–52]. Building on these accounts, the present integration predicts that comparable amounts of delegation can have different consequences when different operations are retained, and that comparable assisted output can conceal different levels of engagement and later unaided capability. These are comparative, testable predictions, not claims that the constituent mechanisms are new or that combining them already validates the framework.

The framework developed in this article translates into a cumulative interdisciplinary research agenda. The propositions outlined in Table 3 are not intended as independent claims of equal theoretical importance. Rather, Proposition 1 specifies the central theoretical mechanism; Propositions 2 and 3 specify how interaction design and the distribution of cognitive operations shape cognitive engagement; Propositions 4 and 5 identify key regulatory processes; Proposition 6 defines important contextual boundary conditions; Proposition 7 specifies a multi-method and longitudinal approach to investigating the proposed relationships; and Proposition 8 articulates the framework's predictions regarding potential long-term cognitive outcomes. Together, these propositions provide an integrated conceptual structure that can guide future empirical research.

Testing these propositions requires separating AI-support features from engagement indicators and outcome measures.

P1 proposes that cognitive engagement, assessed independently through the quality of human cognitive processing, partially mediates the relationship between human–AI cognitive work configuration (including offloading and scaffolding) and short- and intermediate-term outcomes, while allowing for direct effects. P8 addresses the potential long-term consequences of repeated AI use for individuals' cognitive capabilities and willingness to perform tasks independently, rather than outcomes observed during a single AI-assisted task. Section 5.3 illustrates how these distinctions can guide an empirical design.

Table 3. Major Propositions that Result from the Interdisciplinary Framework.

Theoretical Component	ID	Proposed Relationship
Core theoretical mechanism	P1	Cognitive engagement is proposed to partly mediate the relationship between human–AI cognitive work configuration (including cognitive offloading and scaffolding) and short- and intermediate-term outcomes; direct effects of cognitive work configuration on these outcomes may also occur.
Interaction design and cognitive work distribution	P2	Answer-first GenAI design may reduce cognitive engagement through the offloading of cognitive operations consequential for understanding, learning, critical evaluation, or independent judgment.
Interaction design and cognitive work distribution	P3	GenAI systems that provide contingent support for attempts, explanation, comparison, critique, or revision are expected to sustain or enhance independently assessed cognitive engagement, conditional on user and task characteristics.
Regulatory processes	P4	Trust calibration shapes the distribution of cognitive operations: calibrated trust promotes selective delegation and verification, whereas overtrust increases delegation without sufficient human evaluation.
Regulatory processes	P5	Metacognitive regulation shapes whether reductions in cognitive load preserve or diminish meaningful cognitive engagement through monitoring of understanding, confidence, and uncertainty.
Contextual boundary conditions	P6	The effects of GenAI support are contingent on user, task, and system characteristics, as well as social/organizational context, including expertise, accountability, incentives, coordination demands, goals, time pressure, task complexity, and novelty.
Multi-method and longitudinal investigation	P7	Different patterns of cognitive work distribution and cognitive engagement are expected to produce distinguishable behavioral, self-report, and neurophysiological profiles across time; convergent and longitudinal evidence may distinguish reductions in cognitive burden from reductions in meaningful cognitive engagement.
Potential long-term cognitive outcomes	P8	Across repeated GenAI-supported interactions, preserving meaningful cognitive engagement is expected to contribute to cognitive sustainability, whereas repeated delegation with limited engagement may increase cognitive dependence. Reduced unaided ability and reduced willingness to work independently are distinct outcomes to assess separately.

5.2. Evidence Status and Boundary Conditions

The arguments advanced in this paper should be interpreted with deliberate caution. Some claims are grounded in established evidence from GenAI and related studies, whereas others constitute theoretically derived inferences, and claims about longer-term effects necessarily remain more speculative. Explicitly differentiating among these levels of evidential support sharpens the theoretical argument, avoids overstating what is currently known, and delineates priorities for future empirical inquiry. Table 4 therefore classifies the framework's main claims according to their current evidential status, distinguishing

established evidence, emerging GenAI evidence, theoretically grounded inference, and propositions that remain part of the future research agenda.

Table 4. Evidence status of the main mechanisms outlined in the interdisciplinary framework.

Evidence Level	Claims
Established evidence	External tools can reduce internal demands and alter immediate performance and memory [1,20–22]. Active processing, self-explanation, and retrieval can support learning [27,41–43]. Trust calibration informs appropriate reliance [46,48], and contingent, faded scaffolding supports learning [38,45]. These findings come from heterogeneous tasks and do not validate the full GenAI framework.
Emerging GenAI evidence	Productivity benefits vary across tasks and settings [8–10]. Confidence in AI is associated with self-reported critical-thinking effort [11]; task-specific EEG and ownership findings remain preliminary [12], preprint. Unguarded tutoring can impair subsequent unaided performance, with safeguards mitigating that penalty [23]. Programming interfaces show design-dependent effects on engagement and calibration [24].
Plausible theoretical inference	P1–P6 propose links among regulation, operation-level allocation, cognitive engagement, and outcomes. Offloading, engagement, scaffolding, metacognition, and reliance accounts motivate these links [1,27,32–35,38,46,52], but do not establish their proposed combination or mediation. Social-context extensions draw on human-group evidence [53–55], not direct GenAI validation.
Speculation/research agenda	P7–P8 propose convergent, longitudinal tests of persistent cognitive sustainability and dependence. Cognitive-sustainability accounts provide a framing lens [18,19], and learning/offloading studies motivate follow-up [1,22,23]; these sources do not establish durable GenAI-induced cognitive or neural change.

Boundary conditions are central to the framework. Offloading may be appropriate when the operation is routine, already mastered, or easy to verify, although these conditions do not guarantee the absence of longer-term costs. Scaffolding is not beneficial when it adds irrelevant friction, overwhelms users, prevents fluent performance after mastery, or fails to fade. Cognitive sustainability therefore requires adaptive design and selective delegation of cognitive operations rather than a blanket preference for more or less automation.

5.3. Illustrative Empirical Test

An illustrative analytical-report task makes the proposed tests concrete (Table 5). Participants would interpret the same evidence pack and prepare a report under answer-first AI, scaffolded AI, or no-AI conditions. Prior knowledge would be assessed before random assignment, and task materials and time allowance standardized. Support features, human processing, effort, and assisted versus unaided outcomes would be assessed separately. This is a hypothetical design, not a study conducted for this article; the three-condition comparison tests packages of support, not the isolated effect of each design feature.

For P1, an analysis would prespecify the indirect path through engagement and the residual direct path, adjust for baseline capability, and assess the reliability and distinctness of the measures. With adequate precision and reliable measures, an indirect-effect interval excluding effects of a prespecified meaningful magnitude would challenge the proposed mediation. A null total condition effect alone would not exclude an indirect pathway, because direct and indirect effects may differ in direction. Randomizing the assistance condition alone does not eliminate mediator-outcome confounding; causal mediation would require

additional assumptions, sensitivity analyses, or a more targeted manipulation. Operation-specific follow-up experiments would be needed to distinguish the effects of how cognitive operations are distributed between humans and AI from the effects of how these operations are supported by AI.

Table 5. Illustrative test of the framework using an AI-assisted analytical report. This hypothetical design reports no new data or results.

Element	Illustrative Analytical-Report Design
Task and allocation	Use the same evidence pack. Log who formulates the question, selects and checks evidence, evaluates alternatives, drafts, and revises. Comparable amounts of delegation may involve different operations.
Three conditions	Answer-first: AI drafts and the user reviews. Scaffolded: the user attempts; AI provides contingent hints, critique, and feedback, with support adapted to prior knowledge and faded as competence grows. No AI: the user completes the same task independently.
Engagement quality	At matched checkpoints, independent raters, blinded where feasible, score human reasoning, source checking, integration, and substantive revision using a prespecified rubric. Collect comparable process evidence in every condition; do not score required actions or AI text as human engagement.
Effort and regulation	Assess perceived mental effort separately [28]. Measure confidence calibration and verification choices as regulatory indicators, not as substitutes for engagement quality. Record time on task rather than equating longer time with deeper processing.
Immediate outcomes	Separate raters assess report accuracy, evidential justification, and quality; record completion time. Do not use the assisted report score as the engagement measure.
Delayed unaided outcomes	At a prespecified delay, assess recall, explanation of the underlying reasoning, and transfer to a new report without AI. Control baseline knowledge and use outcomes distinct from the engagement rubric.
Predictions and scope	P1–P3: similar assisted output may coexist with different engagement and delayed unaided performance. Prespecify moderation by prior knowledge (P6). These comparisons do not establish P8 or test every regulatory proposition.
Repeated-use extension	For P8, repeat exposure and follow-up. Assess unaided ability on required tasks and willingness through a separate optional unaided task under standardized incentives and time costs; do not infer inability from choosing AI.

Testing P8 would require repeated exposure to AI assistance and follow-up assessments at prespecified intervals. Changes in the ability and willingness to perform tasks independently without AI should be assessed separately, while accounting for baseline differences. The analysis should also distinguish between maintaining and improving cognitive capabilities, as both are relevant to cognitive sustainability. If sustained differences in cognitive engagement were not followed by meaningful differences in either unaided ability or willingness, with equivalence established separately for both outcomes within

prespecified bounds, this would challenge the proposed long-term pathway. However, a single nonsignificant result or one delayed assessment would not provide sufficient evidence. Neurophysiological differences alone would not establish cognitive sustainability without corresponding evidence of lasting effects on human cognitive functioning.

A NeuroIS extension of this illustrative design could complement behavioral and self-report measures with neurophysiological recordings across all three experimental conditions (answer-first AI, scaffolded AI, and no AI). For example, EEG or fNIRS could be used to investigate neural activity associated with task processing and cognitive control, while electrodermal activity and cardiovascular measures could provide complementary indicators of physiological arousal. Recordings could be collected during comparable task stages, including evidence evaluation, verification, and revision, while accounting for differences in task demands and interaction patterns across conditions. Comparing neurophysiological responses across the three conditions could help identify whether different configurations of human–AI cognitive work are associated with distinct patterns of neural and physiological activity. Importantly, such signals should not be interpreted as direct or unique measures of cognitive engagement, and their explanatory value beyond behavioral and self-report indicators would need to be established empirically. Neurophysiological measurements represent a valuable methodological extension that can provide unique insights into the neural and physiological processes underlying human–AI interaction, while remaining complementary rather than indispensable to testing the framework. Crucially, establishing long-term cognitive sustainability requires longitudinal evidence demonstrating the preservation or improvement of human cognitive capabilities.

6. Limitations and Conclusions

The framework has several limitations. First, it is conceptual and requires empirical testing across tasks, user groups, and organizational contexts. Second, current evidence on long-term neurocognitive consequences of GenAI use remains limited, so claims about durable neural adaptation and persistent changes in cognitive functioning should remain hypotheses requiring longitudinal investigation. Third, although social and organizational conditions are included as contextual influences, the target outcomes remain primarily individual; a multilevel theory of team and organizational capability would require further development. The source selection is theory-guided and non-exhaustive, so alternative constructs, relationships, and interpretations may be underrepresented.

The framework's contribution is the integration of operation-level cognitive work distribution, regulation, and processing quality with distinct assisted and unaided outcomes. It generates comparative predictions about what is delegated and what the user continues to process, while treating long-term cognitive sustainability and dependence as hypotheses. Its value therefore depends on whether these distinctions explain outcomes beyond existing accounts and survive the tests outlined in Section 5.3.

NeuroIS contributes complementary process measurements to this research agenda; however, it is not a prerequisite for testing every proposition. Behavioral evidence and longitudinal assessment remain essential for establishing whether human cognitive capabilities are maintained or improved over time. Accordingly, neurophysiological measures should complement, rather than replace, behavioral and self-report measures.

The article also repositions trust. Within this framework, trust is assigned a regulatory role rather than assumed to be the principal mediator. Trust calibration shapes reliance and verification behavior [46,63,64], consistent with broader calls for transdisciplinary trust research in the artificial intelligence era [65]; cognitive engagement captures the depth, persistence, and strategic quality of users' cognitive processing during AI-supported task

performance. This distinction assigns trust a clearly specified regulatory role without treating it as the principal explanatory mechanism.

Moreover, the article gives cognitive sustainability a precise but modest role. It is not presented as a fully new psychological theory. It is used as an integrative lens for evaluating whether AI-supported practices preserve and develop human capacities for thinking, learning, judgment, and agency. This framing is especially appropriate for sustainability research because sustainability concerns not only environmental resources, but also human, social, cultural, and developmental capacities [18,19,66].

The practical question is consequently which combinations of delegation and support preserve the capabilities required in a particular context, not whether all cognitive operations should remain with the individual.

For the research community—including scholars from NeuroIS, cognitive neuroscience, neuropsychology, and neuroergonomics, among others—this represents a unique and timely opportunity. Researchers and designers can contribute to intelligent and efficient technologies that support cognitively sustainable human–AI interaction. Such technologies should not merely produce better outputs. They should also help preserve human capacities through which people learn, understand, judge, and act responsibly.

Ultimately, the key question is not whether AI can think for us. It is whether AI systems will be designed in ways that allow us to continue thinking for ourselves. GenAI will shape the future of human cognition not simply by what it can do, but by what it still asks us to do. The central design challenge is therefore to build systems that support productivity without undermining meaningful human engagement in the cognitively consequential operations on which sustainable human learning, judgment, and agency depend.

Author Contributions: Conceptualization, R.R.; writing—original draft preparation, R.R. and F.K.; writing—review and editing, R.R. and F.K.; project administration, R.R. and F.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: Supported by Johannes Kepler Open Access Publishing Fund and the federal state Upper Austria. During the preparation of this manuscript, the authors used ChatGPT (OpenAI; GPT-5.5 Pro) to support literature identification and synthesis, language editing, structural refinement, drafting of revision suggestions, and development of Figure 1. During the revision process, OpenAI's GPT-6 was additionally used for the same purposes. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Risko, E.F.; Gilbert, S.J. Cognitive offloading. *Trends Cogn. Sci.* **2016**, *20*, 676–688. [CrossRef] [PubMed]
2. Hertel, P.T. Implications of external memory for investigations of mind. *Appl. Cogn. Psychol.* **1993**, *7*, 665–674. [CrossRef]
3. Norman, D.A. Cognitive artifacts. In *Designing Interaction: Psychology at the Human-Computer Interface*; Carroll, J.M., Ed.; Cambridge University Press: Cambridge, UK, 1991; pp. 17–38.
4. Wegner, D.M. Transactive memory: A contemporary analysis of the group mind. In *Theories of Group Behavior*; Mullen, B., Goethals, G.R., Eds.; Springer: New York, NY, USA, 1987; pp. 185–208.
5. Hutchins, E. *Cognition in the Wild*; MIT Press: Cambridge, MA, USA, 1995.
6. Hollan, J.; Hutchins, E.; Kirsh, D. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.* **2000**, *7*, 174–196. [CrossRef]

7. Kirsh, D.; Maglio, P. On distinguishing epistemic from pragmatic action. *Cogn. Sci.* **1994**, *18*, 513–549. [CrossRef] [PubMed]
8. Noy, S.; Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **2023**, *381*, 187–192. [CrossRef] [PubMed]
9. Brynjolfsson, E.; Li, D.; Raymond, L. Generative AI at work. *Q. J. Econ.* **2025**, *140*, 889–942. [CrossRef]
10. Dell'Acqua, F.; McFowland, E., III; Mollick, E.R.; Lifshitz, H.; Kellogg, K.C.; Rajendran, S.; Krayer, L.; Candelon, F.; Lakhani, K.R. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organ. Sci.* **2026**, *37*, 403–423. [CrossRef]
11. Lee, H.-P.; Sarkar, A.; Tankelevitch, L.; Drosos, I.; Rintel, S.; Banks, R.; Wilson, N. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 26 April–1 May 2025; p. 1121. [CrossRef]
12. Kosmyna, N.; Hauptmann, E.; Yuan, Y.T.; Situ, J.; Liao, X.-H.; Beresnitzky, A.V.; Braunstein, I.; Maes, P. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv* **2025**, arXiv:2506.08872. [CrossRef]
13. Riedl, R.; Banker, R.D.; Benbasat, I.; Davis, F.D.; Dennis, A.R.; Dimoka, A.; Gefen, D.; Gupta, A.; Ischebeck, A.; Kenning, P.; et al. On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. *Commun. Assoc. Inf. Syst.* **2010**, *27*, 243–264. [CrossRef]
14. Dimoka, A.; Pavlou, P.A.; Davis, F.D. NeuroIS: The potential of cognitive neuroscience for information systems research. *Inf. Syst. Res.* **2011**, *22*, 687–702. [CrossRef]
15. Dimoka, A.; Banker, R.D.; Benbasat, I.; Davis, F.D.; Dennis, A.R.; Gefen, D.; Gupta, A.; Ischebeck, A.; Kenning, P.; Pavlou, P.A.; et al. On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Q.* **2012**, *36*, 679–702. [CrossRef]
16. Tams, S.; Hill, K.; Ortiz de Guinea, A.; Thatcher, J.B.; Grover, V. NeuroIS—Alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *J. Assoc. Inf. Syst.* **2014**, *15*, 723–753. [CrossRef]
17. Riedl, R.; Fischer, T.; Léger, P.-M.; Davis, F.D. A decade of NeuroIS research: Progress, challenges, and future directions. *DATA BASE Adv. Inf. Syst.* **2020**, *51*, 13–54. [CrossRef]
18. Zöldy, M.; Csete, M.S.; Kolozsi, P.P.; Bordás, P.; Török, Á. Cognitive sustainability. *Cogn. Sustain.* **2022**, *1*, 3–9. [CrossRef]
19. Bray, D.A. Conceptualizing information systems and cognitive sustainability in 21st century 'attention' economies (includes syllabus). *SSRN Electron. J.* **2007**, 1–16. [CrossRef]
20. Sparrow, B.; Liu, J.; Wegner, D.M. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science* **2011**, *333*, 776–778. [CrossRef] [PubMed]
21. Storm, B.C.; Stone, S.M. Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychol. Sci.* **2015**, *26*, 182–188. [CrossRef] [PubMed]
22. Dahmani, L.; Bohbot, V.D. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Sci. Rep.* **2020**, *10*, 6310. [CrossRef] [PubMed]
23. Bastani, H.; Bastani, O.; Sungu, A.; Ge, H.; Kabakcı, Ö.; Mariman, R. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2422633122. [CrossRef] [PubMed]
24. Kazemitabaar, M.; Huang, O.; Suh, S.; Henley, A.Z.; Grossman, T. Exploring the design space of cognitive engagement techniques with AI-generated code for enhanced learning. In Proceedings of the 30th International Conference on Intelligent User Interfaces, Cagliari, Italy, 24–27 March 2025; pp. 695–714. [CrossRef]
25. Corno, L.; Mandinach, E.B. The role of cognitive engagement in classroom learning and motivation. *Educ. Psychol.* **1983**, *18*, 88–108. [CrossRef]
26. Fredricks, J.A.; Blumenfeld, P.C.; Paris, A.H. School engagement: Potential of the concept, state of the evidence. *Rev. Educ. Res.* **2004**, *74*, 59–109. [CrossRef]
27. Chi, M.T.H.; Wylie, R. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educ. Psychol.* **2014**, *49*, 219–243. [CrossRef]
28. Paas, F.; Tuovinen, J.E.; Tabbers, H.; Van Gerven, P.W.M. Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **2003**, *38*, 63–71. [CrossRef] [PubMed]
29. Kahneman, D. *Attention and Effort*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1973.
30. O'Brien, H.L.; Toms, E.G. What is user engagement? A conceptual framework for defining user engagement with technology. *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 938–955. [CrossRef]
31. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*; Harper & Row: New York, NY, USA, 1990.
32. Flavell, J.H. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *Am. Psychol.* **1979**, *34*, 906–911. [CrossRef]
33. Nelson, T.O.; Narens, L. Metamemory: A theoretical framework and new findings. *Psychol. Learn. Motiv.* **1990**, *26*, 125–173. [CrossRef]

34. Zimmerman, B.J. Attaining self-regulation: A social cognitive perspective. In *Handbook of Self-Regulation*; Boekaerts, M., Pintrich, P.R., Zeidner, M., Eds.; Academic Press: San Diego, CA, USA, 2000; pp. 13–39.
35. Gajos, K.Z.; Mamykina, L. Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In Proceedings of the 27th International Conference on Intelligent User Interfaces, Helsinki, Finland, 22–25 March 2022; pp. 794–806. [CrossRef]
36. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978.
37. Wood, D.; Bruner, J.S.; Ross, G. The role of tutoring in problem solving. *J. Child Psychol. Psychiatry* **1976**, *17*, 89–100. [CrossRef] [PubMed]
38. van de Pol, J.; Volman, M.; Beishuizen, J. Scaffolding in teacher-student interaction: A decade of research. *Educ. Psychol. Rev.* **2010**, *22*, 271–296. [CrossRef]
39. Reiser, B.J. Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *J. Learn. Sci.* **2004**, *13*, 273–304. [CrossRef] [PubMed]
40. Quintana, C.; Reiser, B.J.; Davis, E.A.; Krajcik, J.; Fretz, E.; Duncan, R.G.; Kyza, E.; Edelson, D.; Soloway, E. A scaffolding design framework for software to support science inquiry. *J. Learn. Sci.* **2004**, *13*, 337–386. [CrossRef] [PubMed]
41. Chi, M.T.H.; de Leeuw, N.; Chiu, M.-H.; LaVancher, C. Eliciting self-explanations improves understanding. *Cogn. Sci.* **1994**, *18*, 439–477. [CrossRef] [PubMed]
42. Roediger, H.L.; Karpicke, J.D. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.* **2006**, *17*, 249–255. [CrossRef] [PubMed]
43. Karpicke, J.D.; Blunt, J.R. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* **2011**, *331*, 772–775. [CrossRef] [PubMed]
44. Kapur, M. Productive failure. *Cogn. Instr.* **2008**, *26*, 379–424. [CrossRef]
45. Belland, B.R.; Walker, A.E.; Kim, N.J.; Lefler, M. Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Rev. Educ. Res.* **2017**, *87*, 309–344. [CrossRef] [PubMed]
46. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef] [PubMed]
47. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* **1997**, *39*, 230–253. [CrossRef]
48. Hoff, K.A.; Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* **2015**, *57*, 407–434. [CrossRef] [PubMed]
49. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
50. Buçinca, Z.; Malaya, M.B.; Gajos, K.Z. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. In *Proceedings of the ACM on Human-Computer Interaction*; 5(CSCW1); ACM: New York, NY, USA, 2021; p. 188. [CrossRef]
51. Bansal, G.; Wu, T.; Zhou, J.; Fok, R.; Nushi, B.; Kamar, E.; Ribeiro, M.T.; Weld, D.S. Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 8–13 May 2021; pp. 1–16. [CrossRef]
52. Tankelevitch, L.; Kewenig, V.; Simkute, A.; Scott, A.E.; Sarkar, A.; Sellen, A.; Rintel, S. The metacognitive demands and opportunities of generative AI. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–24. [CrossRef]
53. Lewis, K. Measuring transactive memory systems in the field: Scale development and validation. *J. Appl. Psychol.* **2003**, *88*, 587–604. [CrossRef] [PubMed]
54. Bond, C.F., Jr.; Titus, L.J. Social facilitation: A meta-analysis of 241 studies. *Psychol. Bull.* **1983**, *94*, 265–292. [CrossRef]
55. Lerner, J.S.; Tetlock, P.E. Accounting for the effects of accountability. *Psychol. Bull.* **1999**, *125*, 255–275. [CrossRef] [PubMed]
56. Friedman, N.P.; Robbins, T.W. The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology* **2022**, *47*, 72–89. [CrossRef] [PubMed]
57. Somon, B.; Campagne, A.; Delorme, A.; Berberian, B. Performance monitoring applied to system supervision. *Front. Hum. Neurosci.* **2017**, *11*, 360. [CrossRef] [PubMed]
58. vom Brocke, J.; Riedl, R.; Léger, P.-M. Application strategies for neuroscience in information systems design science research. *J. Comput. Inf. Syst.* **2013**, *53*, 1–13. [CrossRef]
59. Shneiderman, B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int. J. Hum.-Comput. Interact.* **2020**, *36*, 495–504. [CrossRef]
60. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K.; et al. Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. 3. [CrossRef]
61. Freeman, S.; Eddy, S.L.; McDonough, M.; Smith, M.K.; Okoroafor, N.; Jordt, H.; Wenderoth, M.P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8410–8415. [CrossRef] [PubMed]

62. Riedl, R. Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electron. Mark.* **2022**, *32*, 2021–2051. [CrossRef]
63. Schaefer, K.E.; Chen, J.Y.C.; Szalma, J.L.; Hancock, P.A. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Hum. Factors* **2016**, *58*, 377–400. [CrossRef] [PubMed]
64. Schemmer, M.; Kühl, N.; Benz, C.; Bartos, A.; Satzger, G. Appropriate reliance on AI advice: Conceptualization and the effect of explanations. In Proceedings of the 28th International Conference on Intelligent User Interfaces, Sydney, NSW, Australia, 27–31 March 2023; pp. 410–422. [CrossRef]
65. Krueger, F.; Riedl, R.; Bartz, J.A.; Cook, K.S.; Gefen, D.; Hancock, P.A.; Jarvenpaa, S.L.; Krabbendam, L.; Lee, M.R.; Mayer, R.C.; et al. A call for transdisciplinary trust research in the artificial intelligence era. *Humanit. Soc. Sci. Commun.* **2025**, *12*, 1124. [CrossRef]
66. Garrigos-Simon, F.J.; Botella-Carrubi, M.D.; Gonzalez-Cruz, T.F. Social capital, human capital, and sustainability: A bibliometric and visualization analysis. *Sustainability* **2018**, *10*, 4751. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.